University of Namur

THESIS / THÈSE

DOCTOR OF ECONOMICS AND BUSINESS MANAGEMENT

Essays in development economics and applied econometrics

Gelade, Wouter

Award date:
2016

Awarding institution:
University of Namur

University of Namur

Essays in Development Economics and Applied Econometrics

Wouter Gelade

Supervisors:
Catherine Guirkinger (University of Namur)
Jean-Marie Baland (University of Namur)

Other members of the jury:
Guilhem Cassan (University of Namur)
Xavier Giné (World Bank)
William Pariente (Université Catholique de Louvain)

President of the Jury:
Mathias Hungerbûhler (University of Namur)

Rempart de la Vierge, 13
B - 5000 Namur (Belgique)

Imprimé en Belgique
ISBN : 978-2-87037-931-8
Dépôt légal: D/2016/1881/20

Contents

Acknowledgements

Many people contributed directly or indirectly to this thesis. Let me start by thanking the people who suffered most for its realization. Thanks Assan, Ayoro, Batiobo, Foro, Idrissa, Jules, Florence, Gautier, Korotimi, Ira, Marthe, Mathieu, Odette, Ibra, Lassana, Togo, Rasmata and, especially, the invaluable Karim Paré for your never ending dedication to collecting data of the highest quality and for being great colleagues on the field.

My advisors Catherine Guirkinger and Jean-Marie Baland played an important role in the realization of this thesis. I have very much enjoyed working with both of you. I especially appreciated that your doors were always open and the general modesty in your ways of working.

Finally, many of my colleagues contributed directly or indirectly to the work in this thesis. Thanks to Vincenzo Verardi, Quentin Stoeffler, Elena Serfilippi, Michael Carter, Ombeline De Bock and Jean-Philippe Platteau for intellectually stimulating and fruitful collaborations. Thanks to my colleagues at CRED, and especially to my office mates through the years, for making the stay in Namur very pleasant.

Merci!

Introduction

This thesis gathers papers on different subjects. In **Part I**, we study several topics in the field of development economics. The first chapter deals with the impacts of fairtrade from a theoretical point of view. The second chapter is an empirical contribution on the role of external monitoring in joint liability credit groups, while the third chapter is a literature review on the demand for micro-insurance.

More concretely, in **Chapter 1**, titled "'Made in Dignity': the redistributive impact of Fair Trade" and co-authored with Jean-Marie Baland and Cedric Duprez, we develop a model of North-South trade to investigate the impact of Fair Trade. In the absence of a label, Southern producers are exploited by monopsonisitic intermediaries who export to Northern markets. The Fair Trade label certifies the adoption of high labour standards and the payment of fair prices to producers in the South. We first show that the label is never Pareto-improving: the welfare of unlabeled producers in the South falls if and only if the welfare of Northern consumers increases. This is more likely to occur when the label only requires a price premium to be paid to producers or when it certifies alternative production practices that do not entail too large productivity losses. An expansion of Fair Trade tends to exacerbate those effects. We also show that the consequences of fair trade on equilibrium prices are systematically reduced in environments where traders enjoy more market power.

In **Chapter 2**, titled "'Let the punishment be proportionate to the offense': External monitoring and internal sanctioning in joint-liability credit groups" and co-authored with Catherine Guirkinger, we study joint-liability credit groups. Models of joint-liability credit typically involve social sanctions and peer-monitoring by group members as means to overcome moral-hazard problems. Interestingly, in practice joint-liability credit often also involves monitoring by the lender (which is usually depicted as prohibitively expensive in the theory literature). To investigate the role of external monitoring, we conduct an experiment in which we increased external monitoring in existing joint-liability credit groups for agricultural credit in Burkina Faso where the key incentive problem is ex-ante moral hazard. We find that external monitoring crowds out internal monitoring by the group's leaders. Because of this substitution, the overall probability to be monitored is not affected. Despite this, sanctioning is affected. The group moves towards more efficient state-dependent sanctioning. Furthermore, purely internal sanctions are substituted for sanctions that also involve the external agent. This suggests that internal and external monitoring are not perfect substitutes. We argue that a key difference lies in how effectively information from monitoring can be used. Accusing group members of wrongdoing is costly in these groups and a judgment by the external agent, based on his own information, can be used more easily to sanction when needed, and only when needed.

In **Chapter 3**, titled "The demand for micro-insurance: A litterature review" and co-authored with Ombeline De Bock and Jean-Philippe Platteau, we review the litterature on the demand for micro-insurance. Micro-insurance has recently received much attention as a promising tool to protect poor individuals from important shocks. Yet, voluntary demand from people has been low, shedding doubt on the viability of micro-insurance as a useful risk-management tool. To better understand this puzzle, the paper reviews both the theoretical and empirical literature on the demand for insurance. While peoples' lack of understanding of insurances does seem to limit the demand for insurance, several more fundamental factors, such as the price, quality, limited trust in the insurer, and liquidity constraints also seem to have an important role in explaining this puzzle.

Part II contains two papers in applied econometrics. The first paper is a methodological contribution to the way people's expectations can be measured. The second paper discusses the Stata implementation of different robust estimators.

More concretely, in **Chapter 4**, titled "Anchoring when measuring expectations: A methodological experiment in Burkina Faso", we conduct an experiment in the methodology of measuring expectations. We elicit income expectations of prospective migrants in Burkina Faso planning to go and work on a gold mine. To do so, we elicit probabilities to earn different amounts of money, but we randomize these amounts: Some respondents are asked about bigger amounts than others. This seemingly irrelevant methodological change has an important effect on the responses. Median expected incomes are up to 4 times higher when expectations are elicited over big amounts instead of small amounts. We argue that people anchor to the proposed amounts and adjust their responses accordingly. Moreover, this undesirable effect is bigger for people with no previous experience in mining, who likely are less knowledgeable about the income distribution. These results suggest caution is needed when using measured expectations, in particular when respondents do not know the distribution well.

Finally, **Chapter 5**, titled “Time efficient algorithms for robust estimators of location, scale, symmetry and tail heaviness" and co-authored with Vincenzo Verardi and Catherine Vermandele, considers the analysis of the empirical distribution of univariate data, which often includes the computation of some location, scale, skewness and tails heaviness measures. These measures are estimates of specific parameters of the underlying population distribution. Several measures are available but they differ in terms of Gaussian efficiency, robustness with respect to outliers, and meaning in case of asymmetric distributions. We first briefly compare, for each type of parameter (location, scale, skewness, and tail heaviness), the "classical" estimator based on (centered) moments of the empirical distribution, an estimator based on specific quantiles of the distribution, and an estimator built on the basis of pairwise comparisons of the observations. This last one always performs better than the other estimators, namely in terms of robustness, but requires at first sight a heavy computation time of an order of n^2. Fortunately, as explained in Croux and Rousseeuw (1992), the algorithm of Johnson and Mizoguchi (1978) allows to substantially reduce the computation time to an order of $n \log(n)$ and, hence, allows to use the robust estimators based on pairwise comparisons even in very large datasets. This has motivated us to program this algorithm and make it available in Stata: we describe in this paper the sketch of the algorithm and the associated Stata commands. Finally, we illustrate on real data the interest of the computation of these robust estimators by involving them in a normality test of the Jarque-Bera form (Jarque and Bera (1980); Brys et al. (2004a)).

Part I

Development Economics

Chapter 1

'Made in Dignity': the redistributive impact of Fair Trade[1]

[1]This chapter is co-authored with Jean-Marie Baland and Cedric Duprez.

Abstract

In this paper, we develop a model of North-South trade to investigate the impact of Fair Trade. In the absence of a label, Southern producers are exploited by monopsonisitic intermediaries who export to Northern markets. The Fair Trade label certifies the adoption of high labour standards and the payment of fair prices to producers in the South. We first show that the label is never Pareto-improving: the welfare of unlabeled producers in the South falls if and only if the welfare of Northern consumers increases. This is more likely to occur when the label only requires a price premium to be paid to producers or when it certifies alternative production practices that do not entail too large productivity losses. An expansion of Fair Trade tends to exacerbate those effects. We also show that the consequences of fair trade on equilibrium prices are systematically reduced in environments where traders enjoy more market power.

1.1 Introduction

Over the recent years, consumers in the North have expressed an increased concern about the working conditions prevailing in the production of what they import from less developed economies.[2] While this movement probably reflects genuine concern about the welfare of poor producers, fair trade restrictions can also be partly motivated by protectionist motives against 'unfair' competition by countries applying low labour standards.[3] In practice, social labeling programs have developed rapidly over the recent years. The sales of Fairtrade certified products have been growing steadily over the last ten years, with double-digit levels of an annual growth rate. In 2014, Fairtrade certified sales amounted to approximately 5.9 billion Euro worldwide. In 2013, there were 1210 Fairtrade certified producer organizations in 74 producing countries, representing over 1.5 million farmers and workers.[4] Fairtrade products typically include coffee, cocoa, bananas, cane sugar, flowers, tea, cotton, fresh fruits, wine grapes, sports balls, etc. Besides their commercial success, most labeling programs are also actively supported by many international organizations such as ILO, UNICEF and major NGOs (Oxfam, Max Havelaar,...).

Fair trade labels can be seen as an effective way to solve informational asymmetries. In many instances consumers are not well informed on the social and economic conditions under which the good they consume has been produced. Labeling by an independent third party provides them with the appropriate information.[5] Labels are also particularly attractive as they do not rely on coercion but simply provide information to the consumers. The latter are then free to choose, by paying a higher price, to support better production conditions, giving rise to a form of

[2]Various studies show that consumers have a preference for 'fair' products and are willing to pay a premium for fair trade products (e.g Prasad et al., 2004; Hiscox & Smyth, 2011; De Pelsmacker et al., 2005; Loureiro & Lotade, 2005; Basu & Hicks, 2008; Poelman et al., 2008; Tagbata & Sirieix, 2008; Cranfield et al., 2010; Elfenbein & McManus, 2010; Hainmueller et al., 2015; Sirieix et al., 2013).

[3]Numerous proposals have been put forward to incorporate minimum labour standards into international trade rules. See Rodrik (1996); Freeman (1998); Bhagwati (1995) for a discussion on the pertinence of imposing labour standards, in line with the debates on the WTO. See also Maskus (1997); Fischer & Serra (2000); Fung et al. (2001); Brown (1999) for more details on labour standards and international trade.

[4]See www.fairtrade.net , Fairtrade International: "Annual report 2014, 2015" and "Monitoring the scope and benefits of fairtrade, Sixth edition, 2014".

[5]Since Akerlof (1970), market failures due to the lack of information on product quality are well known. Labour standards in the production process is a hidden characteristic of goods which is not revealed to consumers even after consumption, a 'credence' characteristic (Nelson, 1970; Darby & Karni, 1973).

'democracy by the consumers'. One therefore expects labeling programs to improve consumer welfare.[6] The higher prices paid also reward complying producers.[7] Labels can therefore be viewed as a tool in the hands of Southern producers that enables them to price discriminate between different types of consumers.[8] A priori, we can therefore expect social labeling to improve the welfare of both Northern consumers and Southern producers.

This is precisely the question addressed in this paper. To this purpose, we set up a simple North-South trade model and analyze the impact of the introduction of a fair trade label in the export production of the South. In the absence of the label, producers in the South sell their output to competing monopsonist traders who have exclusive access to markets. Their market power is modeled as arising from market frictions: each producer has idiosyncratic preferences over the existing traders, who exploit these preferences by under-pricing the output they purchase. If labeled, a trader pays to producers a higher price and guarantees improved production conditions. We also assume that (a) all consumers in the North are willing to pay a price premium for labeled goods, and (b) the label is perfectly implemented and monitored. Taken together, these assumptions tend to bias the results of the model in favor of a positive impact of labeling. Given the relatively limited scope of fair trade in practice, we focus on situations under which the Northern market is not fully covered by labeled goods, so that some Northern consumers also consume unlabeled goods.

We first show that fair trade cannot be Pareto-improving as it always generates losers among producers or consumers. The welfare of unlabeled workers in the South increases if and only if the welfare of Northern consumers decreases. The intuition behind this result is as follows: if the equilibrium price of unlabeled goods rises, Southern workers in the unlabeled sector are better off but consumers in the North are worse off since, in equilibrium, they are indifferent between consuming the high price labeled good and the low price unlabeled good. (The reverse holds when the unlabeled price falls.) Unlabeled prices increase when the fair trade label certifies working conditions that reduce substantially labour productivity, labour hours or the effort levels of the labeled producers. Finally, we show that the consequences of fair trade on equilibrium prices

[6]See e.g. Zago & Pick (2004); Baksi & Bose (2007); Roe & Sheldon (2007), and Bonroy & Constantatos (2008).

[7]Unlike green or eco-friendly labeling, social labels seek first to directly benefit producers instead of promoting a particular public good such as the environment.

[8]From the firm's point of view, a label raising the demand for labeled goods can be viewed as a form of informative advertising.

are systematically reduced in environments where traders enjoy more market power.

So far the literature has essentially proposed partial equilibrium analyses of fair trade, pointing to the beneficial implications for qualifying producers by reducing the intermediaries' market power (Baumann et al., 2012; Podhorsky, 2015). These benefits may however get dissipated under free entry, as argued by De Janvry et al. (2012) in an interesting empirical illustration from coffee cooperatives in Central America. Some authors have also stressed that some producers in the South may directly suffer from the introduction of Fair Trade: "Ethical trading in Bangladesh has both positive and negative consequences, (...). Working conditions have improved in compliant factories, but workers in non-compliant firms are worse-off." (Murshid et al. 2003, see also Valkila & Nygren 2010; Dragusanu & Nunn 2014; Jaffee 2008). In the present paper, we investigate the properties of fair trade as an instrument to reduce the intermediaries' market power in the South and focus on consequences in terms of welfare. The general equilibrium perspective allows us to also analyze more satisfactorily the demand for fair trade, as well as to to identify among the different components of fair trade those that are more conducive to welfare gains for the producers in the South.[9]

The paper proceeds as follows. In Section 2 we present the model. In Section 3, we first characterize the welfare impacts of the label. We then investigate the consequences of an expansion of fair trade as well as of an increase in the monopsony power of intermediaries in the South. Section 4 concludes.

1.2 The model

We consider an economy with two countries, North and South, denoted by N and S respectively. In each country, there is a continuum of measure 1 of identical individuals, who have one unit of time that they supply inelastically on the labour market. We assume complete specialization in production, with the North producing clothes and the South producing food. The production functions are linear, with labour as the only input. Productivity in the North is equal to γ, each worker producing γ units of clothes. We let clothing be the numeraire so that its price is normalized to 1. The income of a worker in the North is then equal to

[9] Some authors also raise doubts about the beneficial impact of a label 'child labor free' label (see e.g. Brown (1999); Davies (2005); Basu et al. (2006); Edmonds (2007); Doepke & Zilibotti (2010)) and Baland & Duprez 2009) In contrast to the present analysis which focusses on eploitative working conditions or pricing practices, being underage is a fixed characteristic of the worker which cannot be changed by the label.

γ. Productivity in the South is equal to 1, with each worker producing one unit of food.

There are potentially two sectors in the South, the labeled and the unlabeled one, respectively denoted by ℓ and u. We let p_ℓ, p_u stand for the price of labeled and unlabeled food respectively. A label on a unit of food certifies that it has been produced under well defined labour standards and fair wages. Monitoring is perfect so that there is no uncertainty associated with the quality of the label.[10]

1.2.1 The North

In the North, individuals consume food and clothing, but also care about the working conditions under which the Southern goods they consume has been produced. The utility function of a Northern consumer is as follows:

$$U_N = (1 + \lambda\mu)\, c_N^\alpha \, (f_N^\ell + f_N^u)^{1-\alpha} \tag{1.1}$$

where $0 \leq \alpha \leq 1$, c_N represents the amount of clothing, f_N^ℓ, the amount of labeled food and f_N^u, the amount of unlabeled food consumed. λ, is a dummy variable which takes the value 1 when consuming labeled food, and 0 otherwise.[11] He thus receives an extra utility benefit $\mu > 0$ when consuming labeled food instead of unlabeled food.

The budget constraint of a Northern household is given by:

$$c_N + f_N^\ell p_\ell + f_N^u p_u = \gamma$$

1.2.2 The South

Southern producers care about the working conditions they face. As consumers, however, they are not concerned about the labour conditions involved in the food they consume.[12] Their utility from consuming and producing goods is as follows:

$$V_S = (1 + \delta\theta)\, c_S^\alpha \, (f_S^\ell + f_S^u)^{1-\alpha} \tag{1.2}$$

[10] The introduction of uncertain quality, while making the analysis more complex, yields essentially similar results as the ones presented in the paper, as long as consumers are ready to pay a premium for labeled - of uncertain quality - over unlabeled food.

[11] Without loss of generality, we implicitly consider that a particular Northern consumer consumes only one type of food so that either $f_N^\ell = 0$ or $f_N^u = 0$.

[12] This assumption is by no way necessary for the validity of the results. It simply allows us to distinguish between concerned and unconcerned consumers without additional notation. The model, and its results, can trivially be extended to the case where some Southern consumers also care about labour standards, while some Northern consumers are indifferent.

where c_S and f_S^j represent respectively the amount of clothes and food of type j, $j = \ell, u$, consumed. When working under high labour standards, the dummy variable δ takes the value 1 and the worker receives a utility benefit of $\theta \geq 0$. δ is equal to 0 otherwize.[13] The two types of food are perfect substitutes, so that, as a consumer, he purchases the least costly variety.

Unlike Northern producers, Southern producers do not sell their production directly on the world markets. Instead, there is a large number N of intermediaries or traders to whom they sell their output. The producer trades with the intermediary he prefers and these traders differ across several dimensions. First, a trader can either be labeled or unlabeled. Producers can produce labeled food only if they trade with a labeled intermediary. Second, different traders can offer different wages. Let $V_{S,i}$ denote the utility derived from producing and consuming goods when trading through intermediary i.

When a Southern producer trades with a particular intermediary i, he also gets an idiosyncratic benefit ϵ_i. His full utility when trading with his preferred trader p is given by:

$$U_S = V_{S,p} + \epsilon_p \tag{1.3}$$

This idiosyncratic benefit ϵ_i is driven by factors such as the distance to the trader or the quality of their personalized relationship or other (unmodelled here) side benefits he draws from selling to this particular trader. The benefit ϵ_i varies across each possible pair of producer and trader, and is drawn from an i.i.d Gumbel distribution[14] with mean zero and standard deviation $d\,(\pi/\sqrt{6})$. Here, d is a measure of dispersion of the ϵ_i: the larger d, the larger the differences in idiosyncratic benefits and the stronger the preference of a producer for a particular intermediary. Because of these benefits, traders enjoy market power over a particular subset of producers.[15]

[13] The utility benefit θ enters the utility function of the Southern producers multiplicatively to mimic the utility benefit Northern consumers get when consuming however fairtrade. However, the results of this paper also hold with an additive utility benefit θ

[14] The Gumbel distribution, also known as the log Weibull distribution, is quite similar to the normal distribution. But unlike the normal it is skewed to the right. The choice for the Gumbel, as opposed to a normal or uniform distribution, is for reasons of tractability. It allows to derive a closed form solution for the proportion of producers who choose a given intermediary (see Equation 1.4), which is a non-trivial problem as it requires comparing the idiosyncratic benefit for a particular intermediary with the idiosyncratic benefits for all other intermediaries.

[15] It is clear that Southern 'producers' can also be seen as workers employed by a particular employer (called here the 'trader'). The analysis of this situation is

1.2.3 Intermediaries in the South

All intermediaries sell food competitively on the world market and southern producers freely choose which trader to sell their production to. As shown in Thisse & Toulemonde (2010), the proportion of producers P_i choosing to trade with intermediary i is given by:

$$P_i = \frac{\exp\left(\frac{V_{S,i}}{d}\right)}{\sum_{j=1}^{N} \exp\left(\frac{V_{S,j}}{d}\right)} \tag{1.4}$$

Each unlabeled trader decides on the price at which he purchases food (which is the wage earned by the workers he trades with) w_u in order to maximize his profits Π_i:

$$\Pi_i = P_i(p_u - w_u)$$

Profits depend on the number of producers the trader attracts when announcing a purchase price w_u, and on the profit generated by each transaction $(p_u - w_u)$. Maximizing profits, the optimal purchase price w_u is given by[16]:

$$w_u = p_u - \frac{d}{V'_{S,u}(w)} = p_u - \frac{d}{\alpha^{\alpha}(1-\alpha)^{1-\alpha}}\, p_u^{1-\alpha}$$

As expected, traders in equilibrium make profits by offering producers a lower price than the market price of food. They are able to do this because producers have idiosyncratic preferences over intermediaries: when an intermediary reduces the price he pays for food, he looses some, but not all, the producers he trades with. The size of this effect is captured by the dispersion of idiosyncratic preferences d. The larger d, the more "attached" producers are to their traders, and the larger the market power the latter can exercise. In equilibrium, a larger d effectively leads to lower prices paid to producers.

Under a fair trade label, a proportion η of traders within the existing set of intermediaries are chosen randomly and given a label.[17] The selected traders have to comply with the fair trade standards. We assume

identical to the one developed here. To avoid confusion, we will stick in the following to the interpretation of the model in terms of producers and traders.

[16] Here, we assume that N is sufficiently large so that the intermediary does not take into account how changes in the price he offers affects the denominator in Equation 1.4.

[17] In other words, the introduction of a label implies that some of the existing intermediaries become labeled intermediaries. Alternatively, one could also consider a label that introduces *additional* labeled intermediaries and where all unlabeled inter-

that η is small enough for the supply of labeled food not to cover the entire market for food in the North. This assumption reflects the fact that most labeling programs in the world are restricted, owing to the limited monitoring capacities of labeling agencies. Thus, FLO, the umbrella body for Fairtrade ensures compliance with Fair Trade standards through a long and strict certification process, which involves a lengthy initial inspection, followed by regular on-site visits. As a result, at the end of 2013, FLO had certified only 1210 producer organizations.

When labeled, a trader can sell food on the world market at the price p_ℓ. Under the label, he has to follow a particular wage rule which requires that he offers a piece rate that is π times higher than the one unlabeled workers receive, w_u.[18] The label also implies costly labour standards on producers: to qualify, a producer has to spend $\sigma \geqslant 0$ units of labour per unit of food produced and $\sigma_c \geqslant 0$ units of clothes.[19] The first type of cost captures the idea that improved labour standards imply higher production costs by resorting to less exploitative modes of production, reducing working hours or spending more resources on workers' health and education. The second type of cost, σ_c, occurs if Northern equipment, goods and expertise are involved in the adoption of improved labour standard (and must be paid for at the going wage rate in the North). As a result the net income earned by labeled producers is given by:

$$w_\ell = (1-\sigma)\pi w_u - \sigma_c$$

In the following, we restrict attention to labels that are beneficial to Southern producers, that is, where the utility of consuming and producing of a labeled producer, $V_{S,\ell}$, is at least as big as that of an unlabeled producer, $V_{S,u}$.[20]

Both unlabeled and labeled intermediaries make profits, and have preferences that are identical to the preferences of Southern workers.

---

mediaries remain unlabeled. However, adding intermediaries provides "free utility" to some labeled producers associated to them because of the idiosyncratic benefits they provide. Nonetheless, except for this effect on some labeled producers' welfare, this model is essentially identical to the one we consider: Effects on unlabeled producers and intermediaries and on Northern consumers do not change.

[18] For example, for coffee FLO requires a price premium of 20 dollar cents per pound with a minimum price of 1.4 dollar. In this paper, we model the fairtrade premium as a price premium, but the results are essentially unchanged when using a minimum price, with the exception of Proposition 5, which we discuss below.

[19] We study costly labels because these costs are an important aspect of fairtrade schemes. However, the results of this paper also hold when the label involves only a wage premium and no costs nor the associated utility benefit θ.

[20] For instance, this is always true when $\sigma_c = 0$ and $(1+\theta)\pi(1-\sigma) \geq 1$

Their utility function is therefore given by:

$$U_K = c_K^{\alpha} \left(f_K^{\ell} + f_K^{u}\right)^{1-\alpha} \tag{1.5}$$

where c_S and f_S^j represent respectively the amount of clothes and food of type j, $j = \ell, u$, consumed. Given the Cobb-Douglas nature of these preferences, the distribution of income between traders and workers does not affect the aggregate demand for each type of good.

1.3 Equilibrium prices and welfare implications

We first describe the equilibrium that prevails before labels are introduced. In the pre-label situation, there are no labeled intermediaries ($\eta = 0$) and no labeled food. The equilibrium price for unlabeled food, p^*, can easily be found by equalizing the supply and the demand for clothes:

$$p^* = (1-\alpha)\frac{\gamma}{\alpha} \tag{1.6}$$

In the labeling equilibrium, a fraction $\eta > 0$ of intermediaries are labeled, which implies that all the producers who choose to sell food to one of them is labeled. As said above, we assume that η is small, so that the supply of labeled food does not cover the entire Northern market. As a result, some consumers in the North consume unlabeled food. The equilibrium prices of labeled and unlabeled food must be such as to leave Northern consumers indifferent between the two types of food:

$$p_\ell = (1+\mu)^{\frac{1}{1-\alpha}}\, p_u \tag{1.7}$$

Again using the market clearing condition for clothing, the equilibrium price of unlabeled food is given by:

$$p_u = \frac{1-\alpha}{\alpha} \frac{\gamma - \eta_S \sigma_c}{1 + \eta_S[(1+\mu)^{1/(1-\alpha)}(1-\sigma) - 1]} \tag{1.8}$$

where η_S is the proportion of labeled producers in the South. While the proportion of labeled intermediaries η is exogenously given, the number of labeled producers is endogenous since each producer chooses which intermediary to trade with. This itself depends on the price of unlabeled food, p_u. Using Equation (1.4), we obtain the equilibrium proportion of labeled producers in the South:

$$\eta_S = \left[1 + \frac{1-\eta}{\eta}\Big/\mathrm{Exp}\left(\frac{V_{S,\ell} - V_{S,u}}{d}\right)\right]^{-1}, \text{ where}$$

$$V_{S,\ell} - V_{S,u} = [A p_u^{\alpha} - d][(1+\theta)(1-\sigma)\pi - 1] - (1+\theta)A\frac{\sigma_C}{p_u^{1-\alpha}}$$

where $A = \alpha^\alpha (1-\alpha)^{1-\alpha}$. If we consider the limit case under which the labeled producers are as well off as the unlabeled ones ($V_{S,\ell} = V_{S,u}$), the proportion of labeled producers, η_S, is exactly equal to the proportion of labeled intermediaries, η: every producer trades with the intermediary that gives him the highest idiosyncratic benefit ϵ_i, and for η producers this happens to be a labeled producer. As the gains from labeling increase, for instance with a larger wage premium π or higher utility gain θ, more producers choose to sell to a labeled trader and the share of labeled producers in the economy increases.

1.3.1 The welfare implications of fair trade

The introduction of the label creates a welfare differential between the unlabeled and labeled producers in the South, the latter being relatively better off. In the North, in equilibrium, consumers must be indifferent between labeled and unlabeled food. Compared to the pre-label situation, Northern consumers are better off with the introduction of a label if and only if the price of unlabeled food, p_u, is smaller than the initial price, p^* (their budget set is strictly larger). However, this is exactly the condition under which the welfare of unlabeled producers in the South falls with the introduction of the label. We therefore have:

Proposition 1 *The label is never Pareto improving, nor Pareto deteriorating. The North is better off iff unlabeled producers in the South are worse off.*

Proof. We have already discussed the fact that the North is better off with the introduction of a label if and only if $p_u < p^*$. The second part of the proof requires that unlabeled workers are worse off if and only if $p_u < p^*$. When the price of unlabeled food falls, the price paid to unlabeled producers by their traders, w_u, also falls, but less than proportionately because the market power of intermediaries is lower when p_u is lower. Moreover, unlabeled workers also consume unlabeled food, which becomes cheaper when p_u falls. In the appendix, we show formally that the net effect of a lower p_u on the welfare of unlabeled workers is negative. ■

This proposition is general and applies to a large set of situations. All we need is that 1) some Northern consumers are indifferent between labeled and unlabeled food such that changes in p_u translate into changes in the utility of Northern consumers and 2) that $w_u = f(p_u)$ with $f' > 0$ such that these changes in p_u have the opposite effect on the utility of unlabeled producers.

We now investigate the conditions under which unlabeled producers gain or loose with the introduction of a label. We have:

Proposition 2 *With the introduction of the label, unlabeled workers are better off and Northern consumers are worse off iff*

$$\frac{\sigma_c}{\gamma} + (1+\mu)^{1/(1-\alpha)}(1-\sigma) < 1 \tag{1.9}$$

Proof: see Appendix.

With the introduction of a label, unlabeled producers gain if the price for unlabeled food increases. This can occur because labeling involves some costs in terms of food or reduced productivity, σ, which lowers the total supply of food on the markets. The introduction of a label also reduces the demand for food, as each consumer of labeled food in the North consumes less units of labeled food owing to their higher price. This is due to the Cobb-Douglas utility functions, under which Northern consumers of labeled food spend the same amount on food as before, but pay a higher price. As a result, the quantity of food they purchase is lower.

At the pre-label price, an excess demand for unlabeled food appears when the former effect dominates the latter and the price of unlabelled food increases. In condition 1.9, the term $(1+\mu)^{1/(1-\alpha)}$ measures the fall in the quantity of labeled food demanded while $(1-\sigma)$ measures the decrease in the net supply of food by labeled producers. If the latter is large (and σ_c is small enough), the condition is satisfied and the price of unlabeled food increases. Proposition 2 also shows that unlabeled workers are more likely to loose when σ_c is high. This is due to the fact that costs in terms of clothes convert demand for labeled food into demand for clothes (Northern consumers, by consuming labelled food, indirectly 'consume' more clothing through these costs), making an excess supply of unlabelled food more likely.

In general however, condition 1.9 is relatively restrictive. To see this, let us restrict attention to labels that involve an *effective transfer* from Northern consumers to Southern producers. This is the case when the price premium paid in the North for a labeled producer's output is bigger than the costs involved to implement the label:

$$(p_\ell - p_u)(1-\sigma) \geq \sigma p_u + \sigma_C \tag{1.10}$$

In other words, under a label with effective transfers, the price premium paid by Northern consumers is not fully consumed by costs and can (partially) be passed on to labeled producers in terms of higher wages.

From Conditions 1.7 and 1.10, it follows that a necessary condition for a label to involve effective transfers is that:

$$(1+\mu)^{1/(1-\alpha)}(1-\sigma) \geq 1 \tag{1.11}$$

Comparing this condition with Condition 1.9 it follows that unlabeled producers always loose in this situation. Unlabeled intermediaries also loose:

Proposition 3 *With the introduction of a label with effective transfers, unlabeled workers are always worse off and profits of unlabeled intermediaries decrease.*

Proof: see Appendix.

Note that a label that only involves a pure price premium and no costs always involves effective transfers, and so unlabeled producers and intermediaries are worse off under such a label. There are however two reasons why a label may not involve effective transfers. First, wages of labeled producers may be increased, not by a transfer from Northern consumers, but by a reduction of profits of the intermediaries. Second, the price premium paid by Northern consumers may be fully consumed by implementing better working conditions, reducing labeled producers' wages but improving their working conditions. As labeled producers get a utility benefit θ because of improved working conditions, they can be better off despite a reduction in wages. Nonetheless, we feel that most fair trade scheme aim to implement a label with effective transfers. We will thus focus mainly on such labels while mentioning how results differ when there are no effective transfers.

The impact of the label varies across labeled producers. This is due to the fact that there is a non empty set of labeled producers who, in equilibrium, are just indifferent between selling to a labeled or an unlabeled trader. For these producers, the impact of a label in terms of welfare is identical to that of unlabeled ones. By contrast, producers who were selling to a trader who became labelled are ex post better off than unlabeled producers, since they still sell to their best preferred trader, and enjoy the gains brought by the label. While these 'non switching' producers usually gain with the introduction of a label compared to the pre-label situation, this is not always the case. These producers can loose even with a label with effective transfers. For instance, in the case of a pure price premium, which always has effective transfers, they loose when the price premium is sufficiently small. In general, for a non-switching labeled producer, the welfare gains are larger when the costs of labelling (σ, σ_c) are low and when the gains from improved working conditions (θ)

are large. As expected, a larger price premium (π) also directly benefits these producers at the expense of the labeled traders who make lower profits.

Profits of unlabeled intermediaries decrease because of two reasons. First, they loose on the extensive margin because labeled intermediaries offer better conditions and in this way attract more producers. They will thus have less units of food to sell. Second, they also loose on the intensive margin - profits per unit sold - because prices of unlabeled food fall with the introduction of the label. While, in reaction to this, they reduce wages for unlabeled producers, they can not fully match the reduction in prices because their bargaining power is smaller when prices are lower.

Profits of labeled intermediaries' profits can both increase or decrease following the introduction of the label. They give a wage premium π to farmers and the higher this premium the lower their profits (up to the point where they make no profits). Clearly, with a sufficiently high wage premium, profits will be sufficiently low so that they are lower than their profits before the introduction of the label. On the other hand, there are also several mechanism that can potentially increase labeled intermediaries' profits. By offering better conditions they attract more producers and gain on the extensive margin. The higher the utility benefits from better working conditions, θ, the stronger this effect. Additionally, labeled products command a price premium in the North, determined by the preferences over labeled goods μ, which allows them to gain on the intensive margin. Whether or not labeled intermediaries' gain depends on the relative strength of these effects.

When the label does not provide effective transfers, unlabeled intermediaries' profits can also increase. This can happen when the label causes prices of unlabeled food to increase (when Condition 1.9 is satisfied). In this context, unlabeled intermediaries gain on the intensive margin while still loosing on the extensive margin. If the former effect dominates the latter, profits of unlabeled intermediaries increase.

We now investigate the effects of expanding the fairtrade sector by increasing the number of labelled traders, η. We have:

Proposition 4 *An expansion of a label with effective transfers (i) increases the welfare of Northern consumers, (ii) decreases the welfare of the unlabeled producers who remain ex post unlabeled and (iii) decreases the welfare of producers who were already labeled before the expansion.*

Proof: see Appendix

An expansion of Fair Trade leads to an increase in the number of labeled producers, which magnifies the consequences in terms of welfare of the introduction of a label. Because overall demand for Southern products decreases, both unlabeled producers in the South and producers who were already labeled are worse off. This result however does not imply that expanding fair trade lowers the overall welfare of producers in the South. Indeed, producers who were selling to a trader who became labeled do gain by becoming labeled producers. In general, our simulations over a large set of parameter values show that the overall welfare of producers in the South (measured additively across all producers) can both increase or decrease with an expansion of the label: Expanding a label that has positive effects on aggregate welfare in the South further increases this welfare. Expanding a label with negative effects further decreases welfare in the South.

When a label does not involve effective transfers, these results may be reversed, depending on whether or not unlabeled producers gain with the introduction of the label (Condition 1.9). If this is the case, expanding the label as before magnifies the welfare consequences of the introduction of the label. When unlabeled producers gain, this implies that expanding the label makes both unlabeled and labeled producers better off while further decreasing welfare of Northern consumers.

Finally, we consider the effects of introducing fair trade in economies with different levels of competitiveness. Recall that d, the dispersion in idiosyncratic benefits for the producers, directly measures trade frictions or the lack of competitiveness among traders in the South. When the same label is introduced in a less competitive environment (higher d), fewer producers become labeled because these frictions reduce the mobility of producers across traders.[21] Since fewer producers become labelled in a less competitive environment, the labeled sector is smaller which weakens the effects of a label on the economy.

Proposition 5 *In a less competitive environment, fewer producers switch to the labeled sector. Following the introduction of a label with effective transfers, the reduction in equilibrium prices of unlabeled food is also smaller.*

Proof: see Appendix

[21] Note also that, in our setting, the price premium is proportional to the price offered to unlabeled producers. In a less competitive environment, unlabeled food prices are lower, and so is the price premium that producers obtain from the labeled trader. This is an additional reason why fewer producers become labeled. This would not be true under another type of price premium, such as a minimum price which is independent from the price offered to unlabeled producers. The price rule actually implemented by most fairtrade programs is typically a combination of both systems, with a minimum price when prices are too low and a price premium when prices are high.

This proposition suggests that, in a less competitive environment, unlabeled workers are less negatively affected by the introduction of a label. Indeed, a smaller reduction in the price of food leads to smaller reductions in the wage of unlabeled producers. However, note that wages (and welfare) were already lower in a less competitive environment before the introduction of the label. Hence, the introduction of the label leads to a smaller reduction in wages (in absolute terms) for producers who had lower wages to start with. This complicates welfare comparisons and it is difficult to say whether the introduction of the label has less detrimental effects for unlabeled producers when there is less competition. The same reasoning holds for labeled producers as well who, just like unlabeled producers, are negatively affected by reductions in unlabeled food prices.[22]

We thus show that the negative effects of the label can be smaller when there is less competition. Could this imply that, in a world with labeling, less competion is good for Southern producers? The answer is no. Simulations over a wide range of parameter values show that for all types of producers the first order effect of a reduction in competion (a reduction in producers' wages) always dominates any general equilibrium effects due to the label. Less competition is thus unambiguously bad for Southern producers.

1.4 Concluding comments

In this paper, we develop a model of North-South tradel to investigate the impact of Fair Trade. In the absence of a label, Southern producers are exploited by monopsonisitic intermediaries who export to Northern markets. The Fair Trade label certifies the adoption of high labour standards and the payment of fair prices to producers in the South. We first show that the label is never Pareto-improving: the welfare of unlabeled producers in the South falls if and only if the welfare of Northern consumers increases. This is more likely to occur when the label only requires a price premium to be paid to producers or when it certifies improved production practices that do not entail too large productivity losses. In general, labelled producers benefit from the introduction of Fair Trade, but these gains are lower when Fair Trade includes a larger set of traders and producers. Finally we showed that the effects of Fair

[22] As mentioned above, the price premium of the label is also lower when competition is lower. This is because wages are lower in this setting and the price premium is proportional to these wages. This is a reason why a label has smaller positive effects on labeled producers when there is less competition. However, with a different type of price premium, this would not be true.

Trade on equilibrium prices are systematically reduced in environments where traders enjoy more market power.

Appendix

Proposition 1 *The label is never Pareto improving, nor Pareto deteriorating. The North is better off iff unlabeled producers in the South are worse off.*

Proof (Continued). It remains to be shown that unlabeled producers are worse off if and only if $p_u < p^*$.

Consider the effect of a change in p_u on unlabeled workers' wages:

$$\frac{\partial w_u}{\partial p_u} = 1 - \frac{(1-\alpha)d}{\alpha^\alpha (1-\alpha)^{1-\alpha}} p_u^{-\alpha} \tag{1.12}$$

Assuming an equilibrium exists, $w_u \geq 0$ must hold and, so, $dp_u^{-\alpha}/\alpha^\alpha(1-\alpha)^{1-\alpha} \leq 1$. As a consequence, $\frac{\partial w_u}{\partial p_u} > 0$ and a decrease in p_u always leads to a decrease in wages for unlabeled producers.

Next, consider the utility of unlabeled workers:

$$U_{S,u} = \alpha^\alpha (1-\alpha)^{1-\alpha} w_u^\alpha (1 - \frac{d}{\alpha^\alpha (1-\alpha)^{1-\alpha}} p_u^{-\alpha}) + \epsilon_i$$

A decrease in p_u leads to a decrease in w_u as well as a decrease of the second term in brackets, and has no other effects. Note that ϵ_i does not change as unlabeled workers do not change intermediaries when the label is introduced. Hence, a decrease in p_u leads to a decrease in welfare for unlabeled workers, and vice versa. ■

Proposition 2 *With the introduction of the label, unlabeled workers are better off and Northern consumers are worse off iff*

$$\frac{\sigma_c}{\gamma} + (1+\mu)^{1/(1-\alpha)}(1-\sigma) < 1$$

Proof. As shown in the Proof of Proposition 1, unlabeled workers are better off and the North is worse off iff prices for unlabeled food rise: $p_u > p^*$. It thus suffices to show that $p_u > p^*$ iff Condition 1.9 in the proposition is satisfied.

To do so, consider the ratio of p_u over p^*:

$$\frac{p_u}{p^*} = \frac{1 - \frac{\eta_S \sigma_c}{\gamma}}{1 + \eta_S[(1+\mu)^{1/(1-\alpha)}(1-\sigma) - 1]}$$

It follows immediately that

$$p_u > p^* \iff \frac{\sigma_c}{\gamma} + (1+\mu)^{1/(1-\alpha)}(1-\sigma) < 1 \tag{1.13}$$

Proposition 3 *With the introduction of a label with effective transfers, unlabeled workers are always worse off and profits of unlabeled intermediaries decrease.*

Proof. As discussed in the text, the first part of the statement, that unlabeled workers are worse off with the introduction of a label with effective transfers, is an immediate corollary of Proposition 2. It thus suffices to show that profits of unlabeled intermediaries decrease with the introduction of a label with effective transfers.

To do so, consider the profits of unlabeled intermediaries:

$$\Pi_u = P_u(p_u - w_u)$$

Profits are a function of two components: 1) The number of producers, P_u that choose the unlabeled intermediary (the extensive margin) and 2) the profits made on every unit sold, $p_u - w_u$ (the intensive margin). We will show that, with the introduction of the label, both components decrease and that profits thus decrease.

For the extensive margin, note that prior to the introduction of the label all intermediaries offer the same conditions and attract the same number of producers. When the label is introduced, the share of producers choosing a given intermediary i is given by Equation 1.4. As one would expect, the share of producers P_i is increasing in $V_{S,i}$, the utility of producing and consuming when trading with intermediary i. Since this utility is higher when trading with labeled than with unlabeled intermediaries ($V_{S,\ell} \geq V_{S,u}$), labeled intermediaries attract more producers than unlabeled intermediaries ($P_\ell \geq P_u$). Since prior to the label they attracted the same number of producers, and since the total number of producers is fixed, this implies that the share of producers attracted by an unlabeled intermediary, P_u, decreases.

As for the intensive margin, we have shown in the proof of Proposition 2 that, with the introduction of the label, prices for unlabeled food p_u fall when Condition 1.9 is not satisfied. When the label provides effective transfers, this is always true and thus p_u decreases. Additionally, Equation 1.12 in the proof of Proposition 1 shows how wages, w_u, change when p_u changes. It follows that:

$$\frac{\partial(p_u - w_u)}{\partial p_u} = 1 - \frac{\partial w_u}{\partial p_u} = \frac{(1-\alpha)d}{\alpha^\alpha (1-\alpha)^{1-\alpha}} p_u^{-\alpha}$$

Since $d \geq 0$, $0 \leq \alpha \leq 1$ and $p_u \geq 0$, this expression is always positive. Hence, the reduction in p_u following the introduction of the label leads to a decrease in $p_u - w_u$, the profits per unit sold by the unlabeled intermediary. ■

Proposition 4 *An expansion of a label with effective transfers (i) increases the welfare of Northern consumers, (ii) decreases the welfare of the unlabeled producers who remain ex post unlabeled and (iii) decreases the welfare of producers who were already labeled before the expansion.*

Proof. Since the discussion in the paper following this proposition goes beyond the proposition, we prove a somewhat more general lemma below. However, Lemma 1 immediately implies the proposition. Indeed, by Condition 1.11 any label with effective transfers satisfies the condition $\frac{\sigma_c}{\gamma} + (1+\mu)^{1/(1-\alpha)}(1-\sigma) > 1$ in the lemma. ■

Lemma 1 *If* $\frac{\sigma_c}{\gamma} + (1+\mu)^{1/(1-\alpha)}(1-\sigma) > 1 (< 1)$, *an expansion of a label (i) increases (decreases) the welfare of Northern consumers, (ii) decreases (increases) the welfare of the unlabeled producers who remain ex post unlabeled and (iii) decreases (increases) the welfare of producers who were already labeled before the expansion.*

Proof. We will show that an increase in the number of traders η leads to a decrease (increase) in unlabeled food prices when $\frac{\sigma_c}{\gamma}+(1+\mu)^{1/(1-\alpha)}(1-\sigma) > 1 (< 1)$. This implies the lemma. Indeed, in the proof of Proposition 1 we have shown that a decrease (increase) in unlabeled food prices (1) increases (decreases) the welfare of Northern consumers and (2) decreases (increases) the welfare of Southern producers, provided that they do not switch trader.[23] Since the proposition involves unlabeled producers who remain ex post unlabeled (case ii) and producers who were already labeled (case iii), this suffices.

It thus remains to be shown that increasing η leads to a decrease in unlabeled food prices when

$$\frac{\sigma_c}{\gamma} + (1+\mu)^{1/(1-\alpha)}(1-\sigma) > 1 \tag{1.14}$$

is satisfied and leads to an increase in p_u when it is not satisfied. Let us first prove the first part, that an expansion of fairtrade leads to a decrease in p_u when Condition 1.14 is satisfied.

To this end, recall that apart from p_u, also the share of labeled workers in the South, η_S, is endogenous. To understand how prices change,

[23] In fact, we have only shown that welfare of unlabeled workers falls iff p_u falls. However, the same proof applies for labeled workers substituting the utility function of unlabeled workers for the one of labeled workers.

we thus also need to consider changes in η_S. A change in η has no immediate effect on p_u, but increases η_S:

$$\frac{\partial p_u}{\partial \eta} = 0 \text{ and } \frac{\partial \eta_S}{\partial \eta} > 0 \tag{1.15}$$

Moreover, under Condition 1.14, the immediate effect of an increase in η_S is a decrease in p_u while an increase in p_u leads to an increase in η_S:

$$\frac{\partial p_u}{\partial \eta_S} < 0 \text{ and } \frac{\partial \eta_S}{\partial p_u} > 0 \tag{1.16}$$

The total effect of an increase in η is the following: It increases η_S, which has a negative effect p_u. The latter has a negative effect on η_S, thus attenuating the original change. Despite this attenuation we will show that the overall effect of an increase in η is an increase in η_S and a decrease in p_u.

To this end, consider the different possibilities following an increase in η: (i) both η_S and p_u increase; (ii) both η_S and p_u decrease; (iii) η_S decreases, p_u increases; and (iiii) η_S increases, p_u decreases. We need to show that (i)-(iii) are impossible and hence (iiii) is the only possibility.

Statements (i) and (ii) are impossible because $\frac{\partial p_u}{\partial \eta_S} < 0$ (Equation 1.16) and, moreover, η has no immediate effect on p_u (Equation 1.15). It is thus impossible that η_S and p_u move in the same direction following a change in η.

Finally, to show that (iii) is impossible, note that $\frac{\partial \eta_S}{\partial \eta} > 0$ and $\frac{\partial \eta_S}{\partial p_u} > 0$. This implies that if both η and p_u increase, so should η_S, which is in contradiction with (iii).

Next, let us show that an increase in η leads to a decrease in p_u when Condition 1.14 is not satisfied. This is easier. Indeed, when Condition 1.14 is not satisfied, changes in p_u and η_S go in the same direction:

$$\frac{\partial p_u}{\partial \eta_S} > 0 \text{ and } \frac{\partial \eta_S}{\partial p_u} > 0$$

It is thus immediate that an increase in η, whose immediate effect is to increase η_S, leads to an increase in both η_S and p_u. ■

Proposition 5 *In a less competitive environment, fewer producers switch to the labeled sector. Following the introduction of a label with effective transfers, the reduction in equilibrium prices of unlabeled food is also smaller.*

Proof. We need to show that, in a less competitive environment (higher d), the introduction of a label with effective transfers leads to fewer producers in the labeled sector (lower η_S) and a smaller reduction in equilibrium prices (higher p_u).

The proof is almost identical to the one of Proposition 4 where we looked at the comparative statics of η. The direct effect of a change in d on p_u and η_S is:

$$\frac{\partial p_u}{\partial d} = 0 \text{ and } \frac{\partial \eta_S}{\partial d} < 0 \tag{1.17}$$

Just like for η, a change in d only has a direct effect on η_S and not on p_u. However, the direct effect of d on η_S is negative, while it was positive for η.

Under a label with effective transfers, the price for unlabeled food, p_u, falls following the introduction of the label. Using the same argument as in the proof of Proposition 4, we can show that an increase in d leads to a decrease in η_S and an increase in p_u. Since the original effect of the label was a decrease in p_u, the reduction in p_u is indeed smaller in a less competitive environment. ■

Chapter 2

'Let the punishment be proportionate to the offense': External monitoring and internal sanctioning in joint-liability credit groups[1]

[1]This chapter is co-authored with Catherine Guirkinger

Abstract

Models of joint-liability credit typically involve social sanctions and peer-monitoring by group members as means to overcome moral-hazard problems. Interestingly, in practice joint-liability credit often also involves monitoring by the lender (which is usually depicted as prohibitively expensive in the theory literature). To investigate the role of external monitoring, we conduct an experiment in which we increased external monitoring in existing joint-liability credit groups for agricultural credit in Burkina Faso where the key incentive problem is ex-ante moral hazard. We find that external monitoring crowds out internal monitoring by the group's leaders. Because of this substitution, the overall probability to be monitored is not affected. Despite this, sanctioning is affected. The group moves towards more efficient state-dependent sanctioning. Furthermore, purely internal sanctions are substituted for sanctions that also involve the external agent. This suggests that internal and external monitoring are not perfect substitutes. We argue that a key difference lies in how effectively information from monitoring can be used. Accusing group members of wrongdoing is costly in these groups and a judgment by the external agent, based on his own information, can be used more easily to sanction when needed, and only when needed.

2.1 Introduction

Microfinance has raised great hopes for poverty alleviation by providing the poor with the capital necessary to undergo profitable projects. One of the main elements of the innovating lending technologies of MFIs is joint-liability, whereby individuals who do not possess the collateral necessary to access a classic credit contract may be offered a joint-liability contract in a credit group. Group members are then jointly responsible for each other's debt. To avoid group default, if one member does not pay back debt, other members have to pay it for him. In the absence of formal collateral, joint-liability lending is supposed to rely on peer-monitoring and social capital to overcome informational problems such as moral-hazard. First, group members should have incentives to monitor each other because they bear the cost of each other's default. Second, monitoring should be relatively cheap for them because they are geographically and socially close to each other. Third, group members can use social sanctions to punish default as they interact with each others in many different spheres. The key assumptions underlying joint-liability are thus that this "internal monitoring" is sufficiently cheap, and that group members can act upon the information they gather to sanction each other when needed (Stiglitz 1990; Ghatak & Guinnane 1999), for instance by making use of "social collateral" (Besley & Coate 1995).

If these assumptions hold, the bank should delegate monitoring to the group instead of engaging in expensive "external monitoring". Yet, in practice, joint-liability schemes often involve active external monitoring. Chowdhury (2005) gives the example of Grameen bank where credit officers attend weekly credit meetings, instead of simply having the group pool the money and directly reimburse him. Likewise, in the joint-liability groups for agricultural credit we study here, external agents actively visit individual farmers' fields to detect moral hazard.

In the literature there is not much attention for the role of external monitoring in joint-liability credit and its interactions with internal monitoring. A notable exception is the theoretical work by Chowdhury (2005), who shows that some expensive external monitoring can be necessary to crowd-in sufficient levels of cheaper internal monitoring. This result is driven by strategic complementarity in monitoring efforts: a group member disciplined by monitoring has greater incentive to monitor his peers because he has more at stake in case they default. To our knowledge, this question has not been examined empirically.[2]

[2]Cason et al. (2012) compare joint-liability with internal monitoring with external monitoring under individual-liability, but they do not consider external monitoring under joint-liability.

In this paper, we explicitly study the effect of external monitoring in joint-liability through an experiment that randomly increases external monitoring in some groups. We study groups of cotton farmers in Burkina Faso who take joint-liability loans to acquire inputs for their cotton production. Farmers have an incentive to deviate these inputs to other crops and so a key problem in these groups is ex-ante moral hazard. Both the group's leaders and external agents already monitor farmers and our experiment thus tests for the effect of an increased intensity of the agent's monitoring.

In contrast with Chowdhury's theoretical prediction, we find that increased external monitoring crowds out internal monitoring by the leaders. Because of this substitution, the overall probability that a farmer is monitored does not seem to be affected by our intervention. Nonetheless, the sanctioning regime changes. The group substitutes purely internal sanctions for sanctions decided in the presence of the agent. Moreover, they move to more efficient state-dependent sanctioning, meaning that they sanction more when defaults are caused by misbehavior and sanction less when they are not.

Perhaps the most important result is that a substitution of internal for external monitoring leads to more state-dependency in sanctioning. While there are multiple explanations for this, we argue, based on qualitative evidence, that the most plausible one is that information collected by the external agent can be used more effectively for sanctioning. There seems to be a high "cost of accusing" group members of wrongdoing, and sanctioning only in case of misbehavior implies incurring this cost. In contrast, when information is collected by the external agent, he can cast an external judgment that can then be used to sanction when needed, and only when needed.

Our results thus relate to the literature on the cost of using peer sanctions. When peer sanctions are costly, they might not be sufficiently effective for joint-liability to address moral hazard (Ghatak & Guinnane, 1999). As an example in another context, it is well understood that the possibility of (costly) sanctioning increases cooperation in a public goods game (Ostrom et al. 1992; Gächter & Fehr 2000), but sanctioning and cooperation is reduced when allowing counter-punishment (Nikiforakis, 2008) or when groups contain many people prone to retaliate (Ones & Putterman, 2007). Our results suggest a way to make peer sanctions more acceptable and less costly: objectifying the sanctions by providing external information.

2.2 Data and experimental design

We collected first-hand data on 890 cotton farmers in 40 villages in the area of Houndé in the South-West of Burkina Faso.[3] These farmers belong to 71 different joint-liability credit groups. In each group we interviewed 13 randomly chosen farmers (or all group members if the group has at most 13 members). Additionally, we interviewed the group's leaders (secretary and president). From the leaders we elicited the list of all defaults (including of non-interviewed farmers) in their group based on their bookkeeping records. Table 2.1 shows some descriptive statistics about the farmers.

We collected two rounds of data (in 2013 and 2014), at the end of the agricultural season just following harvests. Five of the original 890 households were not found during the second round and were not interviewed again. The experiment took place in 2013 and we mainly focus on outcomes of the 2013 campaign.[4]

Our experiment consisted in increasing the level of external monitoring in a random sample of credit groups. Randomization was done at the village level with 20 monitored villages and 20 control villages. In a monitored village all interviewed groups were monitored while in a control village no groups were monitored. In these villages we randomly selected the groups to be interviewed, and the sample consists of 31 monitored groups and 40 control groups. [5]

The monitoring treatment involved an increase in monitoring by the extension agent who is in charge of the group. As we extensively discuss below, extension agents are involved in group monitoring already and our intervention thus serves to increase the intensity of this monitoring. Specifically, in treatment groups the agent was instructed to visit the group every 10 days for three months, from the middle to the end of the campaign. The intervention came as a surprise, both for agents and for farmers. At every visit, he had to visit the fields of two farmers of the group and to record some information (status of the crop, status of the

[3]The data collection was part of a research project, funded by USAID, to evaluate a new insurance for these cotton farmers.

[4]Some information relative to the 2013 campaign - such as the settlement of default - was collected in 2014.

[5]Concretely, in a small village with only one or two groups all groups were interviewed. In a big village (holding at least 3 groups), three groups were chosen at random in control villages and two groups were chosen at random in monitored villages. This explains why there are fewer monitored groups than control groups. Additionally, this implies that treatment is only random conditional on the type of village (big or small). We control in all regressions for whether the group is in a big village.

	Mean	Sd	Median	N
Characteristics household head and Household				
Age household head	43.61573	12.85161	42	890
Nr. years of education household head	1.204494	2.493362	0	890
Household head has some education	.2404494	.4275967	0	890
Nr. years household head	15.5748	11.5828	12	889
Nr. years household head is part of credit group	10.05293	6.35288	10	888
Household head close family of a village leader	.2157303	.4115593	0	890
HH size (nr. people at least 6)	8.505618	5.251651	7	890
Consumption proxy (Progress out of Poverty Index, PPI)	36.66854	12.53419	36	890
Self-sufficient in cereal production (last 3 years)	.6115023	.487695	1	852
Agricultural production				
Cotton area cultivated 2013	3.865618	3.405493	3	890
Cotton yield 2013	826.7994	344.363	800	890
Average cotton yield (2008-2012)	876.7581	255.4646	859.9	851
Cultivated GM cotton 2013	.5505618	.4977166	1	890
Cereal area cultivated 2013	4.512277	3.197328	4	887
Credit for cotton production				
Took credit for cotton production 2013	.994382	.0747844	1	890
Total value of cotton production (in CFA)	808247.4	903720.4	520877.5	890
Cotton area for credit	4.049045	3.682221	3	890
Total credit for cotton production (in CFA)	399963	396861.6	275000	877
Defaulted (individual default)	.1254276	.3313921	0	877
At risk of default (reduction of 1 sd in yield leads to default)	.2827822	.4506084	0	877
Monitoring in credit group				
Agent visited field	.5112613	.5001549	1	888
Agent measured field	.3617978	.4807909	0	890
President visited field	.4208543	.4940066	0	796
President measured field	.1559748	.3630599	0	795
President or agent visited field	.7166247	.4509211	1	794
President or agent measured field	.4402516	.4967298	0	795

Sample sizes differ across variables. Out of the 890 farmers interviewed, 38 were in newly formed households and could not report historical cereal and cotton yields. 13 did not know their credit amount, which is also used to calculate defaults, although they did know their credit area. Information about leader monitoring was obtained from the leaders, who were only asked detailed questions for a random subsample of 796 farmers. All other missing are due to individual farmers not knowing the reply (or not wanting to reply) to a particular question

Table 2.1: Descriptive statistics

field, recommendations given, ...). He could freely choose which farmers to visit, but needed to visit different farmers every time. In control groups no such instructions were given.[6]

The randomization appears to be well-balanced. Since there is no baseline, we compare treatment and control on characteristics that are pre-determined, that is, should not be affected by the treatment. Table 2.2 shows differences on measures of agricultural production (cotton area, cotton yields, cereal area and cotton credit), different characteristics of the household and its head as well as measures of household wealth and food security. In this setting, measures of agricultural production and credit are important and we see no significant differences on these measures. The characteristics of the household heads are also similar across control and treatment villages. The only significant differences we can find are whether the household owns a motorbike and whether the household was self-sufficient in cereal production for the last 3 years, that is, produced enough cereals for its own consumption every year. Because of these differences we also show different measures of wealth and food security and find no other significant differences: Overall, treated farmers appear slightly less food secure, while it is unclear whether treated are wealthier or not. In any case, we control in all regressions for these systematic differences.[7]

2.3 Functioning of cotton credit groups

Group structure

In Burkina Faso, cotton producers are typically organized in groups that receive joint-liability loans, pool their input purchases and jointly sell their cotton to the local parastatal cotton company that enjoys a complete monopsony. These groups consist of 8 to 79 producers (34 on

[6] Note, however, that the agents who monitor treated groups might also be in charge of some control groups. Since the monitoring intervention increases the workload of these agents, they might thus have reduced monitoring in the control groups. Nonetheless, we believe that, if it exists, this effect is small. Agents were typically asked to monitor (on average) 3 groups out of the 30 groups they follow, or 10%. The workload for monitoring 3 groups was about 1 day every 10 days, that is, also 10% of their total time available. While agents do not spend all their time visiting groups, the total time required for the additional monitoring thus seems quite limited. Any reduction in monitoring in the control groups should thus be small.

[7] Note that the variable "Self-sufficient in cereal production" is missing for 38 households. This is because these are newly formed households who have just started their agricultural production and for whom this variable is undefined. We use these households and additionally systematically control for this control being missing, that is, for whether households are new.

	Control	sd	Diff: T - C	se	N
Characteristics household head and households					
Age household head	44.16	(12.72)	-0.437	(1.362)	890
Nr. years of education household head	1.000	(2.379)	0.237	(0.252)	890
Household head has some education	0.210	(0.413)	0.0508	(0.0496)	890
Nr. years household head	16.33	(11.86)	-0.236	(0.897)	889
Nr. years household head is part of credit group	10.05	(6.512)	-0.392	(0.641)	888
HH size (nr. people at least 6)	9.055	(5.422)	-0.828	(0.519)	890
Wealth and food security HH					
HH has television	0.209	(0.517)	0.0297	(0.0441)	890
HH has bed	0.647	(1.458)	-0.0686	(0.136)	890
HH has moto	0.857	(1.195)	-0.194**	(0.0905)	890
HH has house with solid walls	0.105	(0.354)	0.0349	(0.0503)	890
Consumption proxy (PPI index)	37.35	(12.30)	-0.380	(1.480)	890
Self-sufficient in cereal production (last 3 years)	0.632	(0.477)	-0.0936*	(0.0509)	852
Somebody reduced meal last year	0.186	(0.393)	0.0243	(0.0313)	890
Somebody skipped meal last year	0.104	(0.291)	0.0187	(0.0244)	890
Agricultural production					
Cotton area cultivated 2013	3.798	(3.505)	-0.218	(0.511)	890
Average cotton yield (2008-2012)	873.9	(260.9)	-21.70	(31.83)	851
Cultivated GM cotton 2013	0.722	(0.498)	0.00686	(0.105)	890
Cereal area cultivated 2013	4.294	(3.126)	-0.114	(0.362)	887
Credit for cotton production					
Cotton area for credit	3.898	(3.764)	-0.246	(0.548)	890
Total credit for cotton production (in CFA)	409708.9	(424258.6)	-50054.6	(59749.1)	877

Sample sizes differ across variables. Out of the 890 farmers interviewed, 38 were in newly formed households and could not report historical cereal and cotton yields. 13 did not know their credit amount, which is also used to calculate defaults, although they did know their credit area. Information about leader monitoring was obtained from the leaders, who were only asked detailed questions for a random subsample of 796 farmers. All other missing are due to individual farmers not knowing the reply (or not wanting to reply) to a particular question

Table 2.2: Table of balance using pre-determined characteristics

average in our sample) living in the same village. Some villages in our area of study (32%) only count one group, the majority contains at most 3 groups (58%) but other, typically larger villages have multiple groups (up to 12 in our sample).

Groups are headed by a president and a secretary. The president is usually a respected elderly while the secretary is often the most educated member of the group. These leaders are the group representatives and spokesmen in front of the cotton company and they are in charge of the group administration. In particular they supervise the admission of new members, they control bookkeeping and manage the group's loan. They organize at least two plenary meetings per year. During the first meeting they collect and discuss each producer's credit demand in order to prepare the group's collective loan and input demand while during the second meeting they distribute the cotton revenue and settle problems of individual default.

Credit contracts

The group loans are granted by a bank in close relationship with the cotton company. The loans are disbursed exclusively in kind, in the form of cotton inputs delivered by the cotton company (seeds, fertilizers, herbicides and pesticides). Loan sizes are proportional to cotton area: the cotton company offers a standardized package of input per hectare that largely determines loan size, though groups have the possibility to give (some) members less than this standardized package. The loan represents, on average, about 40% of gross cotton revenue.

The group loans are collateralized by the group's future cotton production. The joint-liability clause is strictly enforced since the cotton company pays the group's cotton revenue directly to the group's bank account and the group only has access to the revenue net of the group's debt.

The credit contract with the bank stipulates that group defaults are sanctioned by the group's exclusion from future loans, which would imply that farmers would de facto be excluded from cotton production (except if they are accepted in another group). Group defaults are in fact extremely rare and in practice not immediately sanctioned by exclusion. Defaulting groups are often offered "a second chance" and are carefully monitored by the local agent of the cotton company. For instance, one group in default in the study area was temporarily denied a loan, after which they were required to follow a strict plan to reimburse the outstanding debt.

Individual default and moral hazard

While the group level default rate is very low, there are frequently individual defaults, whereby the value of the cotton produced of one farmer is lower than his share of the group loan.[8] In our sample 12.5% of farmers are in this situation following the 2013 campaign (which was not particularly bad).

The high rate of individual default is in part related to the variability of cotton yields in the context of the Soudano-Sahelian climate characterized by erratic rains and very variable levels of pest infestation. In fact when queried about the causes of all defaults they had dealt with over the last year, group leaders reported that 13% of them were related to either an excess or a deficit of rain. (See Table 2.10 in the Appendix for an overview of the causes of default).

However, the most commonly mentioned reasons for default are related to elements under a farmer's control, namely labour and chemical inputs. Farmers do not have access to credit for their other crops (principally cereals) and have an incentive to use a part of each of their chemical inputs (fertilizers, herbicides and pesticides) on these crops. The inputs obtained through the cotton loan are also sometimes sold on the market, but this is much less pervasive than diverting them to other crops. Diversion of labour, where farmers are suspected to neglect working their cotton field to the benefit of other crops, is also important. In fact, "did not work enough" is cited twice as often (22% of cases) as "did not use enough inputs" as cause of default. Overall, ex-ante moral hazard appears to be a pervasive feature of these cotton loans.[9]

Sanctions

In a situation of individual default, the group total cotton revenue covers the individual defaulter's deficit and it is the responsibility of the group (and in particular of its leaders) to decide how the defaulter reim-

[8]Note that this definition implies that a farmer can default even when he uses another source of income to reimburse the loan. We define defaults in this way because any credit that is reimbursed using non-cotton revenue is still a concern for the group. As we discuss below, having to reimbursing the credit immediately (using another source of income) is generally considered as a sanction and imposing it can create tensions in the group.

[9]There is very limited scope for ex-post moral hazard in this scheme, whereby farmers would fail to reimburse their loan despite having produced sufficient amounts of cotton. They could in principle do this by engaging in pirate sales, but this is complicated by the fact that the cotton company is the only purchaser of cotton. While such pirate sales do exist in the cotton systems of other countries, farmers say it does not happen and leaders do not report it as a cause of default.

	Mean	Sd	Median	N
Defaulters				
Reimbursed immediately	.3489583	.4778869	0	192
Change in credit for cotton production (in CFA)	-49836.24	211133.1	-10200	59
Change in cotton area for credit	-.1984127	1.389882	0	63
Abandoned cotton	.0634921	.2458045	0	63
Non-defaulters				
Change in credit for cotton production (in CFA)	10471.73	256540.5	7240	435
Change in cotton area for credit	.100907	2.242353	0	441
Abandoned cotton	.0612245	.2400139	0	441

Different samples are used for defaulters. The 192 observations come from the leader interview covering all defaults. 63 of these defaulters were surveyed. Both for defaulters and non-defaulters there are fewer observations on credit amount than on credit area because some respondents did not know their exact credit amount

Table 2.3: Consequences of default in control groups

burses his debt towards other group members and to enforce this decision. One sanction (applied in 34% of cases) is to request immediate repayment which the defaulter typically finances by selling cereals or animals. Sometimes the defaulter is given until the next harvest to pay back his debt (without additional interest or financial penalties for the year of delay), in which case group members who paid the deficit are penalized. In some cases only part of the group covers the debt, typically producers who are very close to the defaulter.

Another common sanction is to reduce the farmer's loan size in the year following default. Usually, this is done by reducing the cotton area, though another option is to reduce the loan size per hectare cultivated. Table 2.3 shows that, following the 2013 campaign, the credit area and credit size decreased for defaulter while it slightly increased for other farmers. Decisions on loan size and cotton area are taken during a group meeting to which the agent of the cotton company is taking part. At this meeting members declare which area of cotton they would like to cultivate (and thus finance) and the group either validates or opposes the demand. In practice, it is mainly the leaders and the agent who intervene to limit members' cotton area. In extreme cases, an individual can be permanently or temporarily excluded from the credit group. However, this is a sanction that is used only very rarely. In fact, there is about the same proportion of defaulters and non-defaulters that stopped doing cotton in 2014 in the control, suggesting it was not used as a sanction for default.

Leaders as well as group members' interviews suggest that the management of default is the most challenging aspect of cotton production. Applying sanctions is often difficult and individual defaults generate tensions among group members who typically have close social ties. Agents report that leaders often ask them to intervene at the credit meeting to reduce someone's credit because they cannot do it themselves. Interestingly managers of the cotton company also indicate that individual defaults are a continuous source of concern for them, even when it does not cause group defaults. The reason is that individual defaults within joint-liability credit groups may discourage investment in cotton production as they decrease its profitability (and increase the variability), in particular for "good producers" who are never defaulting.

Information and monitoring

The literature on joint-liability credit suggests that peer monitoring should limit opportunistic behavior such as ex-ante moral hazard. At first sight monitoring may appear particularly easy in the context of cotton groups. First, group members all produce cotton and thus have first-hand knowledge that should help them gauge the state of a crop or detect under-application of inputs. Second, farmers live close to each other and should therefore be able to visit each other's field at low cost. Direct peer-monitoring is however less widespread than we expected. It appears difficult for a regular group member to make a courtesy visit to a peer's field, as those visits are perceived as intrusive. Our information on field visits in the control group in 2014 indicate that only 33% of farmers received a visit of any of their (non-leader) peers in their cotton field.[10]

In contrast, the president and secretary are actively engaged in monitoring: in the control group and in the same year 59% of sampled farmers have received the visit of a group leader. When visiting a field they always inspect the field, and often also measure its area. These two forms of monitoring allow to detect different forms of input diversion. Inspecting the field allows to detect under application of inputs on a given area while measuring can detect a reduction in cotton area, which also frees up inputs to apply on other crops. During interviews leaders indicate that field visits help them dealing with defaulting farmers as it enables

[10] Qualitative information suggests that also sanctioning is mainly the responsibility of the leaders. Nonetheless, there is some peer monitoring and it would be interesting to also investigate the effect of the intervention on peer monitoring and sanctioning. We can not do this because we did not collect data on peer monitoring in 2013, the year in which the intervention was implemented.

them identify the cause of default. However, even leaders do not feel that their visits are welcome. Some of them mentioned that a strong suspicion of default is necessary to justify a field visit. They explained that the visits may offend the farmer who will interpret it as a sign that the leader does not trust him anymore. Other leaders visit every farmer in the group exactly once per campaign to avoid casting blame on anyone in particular.

The cotton company also engages in direct monitoring of individual farmers. Local extension agents of the cotton company (ATC) not only meet with the group leaders but also visit and measure some farmers' fields during the cotton campaign. Overall, over the campaign, extension agents visit on average 37% of group members and measure the visited fields of 24% of farmers, in line with the recommendation of the cotton company. These visits have several objectives. They enable agents to detect cases of credit diversion and also to adapt their global recommendation in terms of pest management to the state of the crop. Cotton agents thus engage in relatively costly monitoring. Managers of the cotton company perceive this follow-up as necessary and feel that in its absence there would be too many internal problems, eventually reducing cotton production. They are in fact steadily increasing the level of monitoring: In the near future they aim at measuring every cotton field every single year and to collect data on individual farmers, rather than groups, to make a more individualized follow-up possible.

In our area of study, agents and group leaders are generally on good terms. Leaders and group members show a lot of respect for the agent, possibly because next to being credit officer he is also an extension agent. While the agent is not a member of the rural community he works in, he lives in a nearby village and leaders often go and talk to him about practical matters or problems in managing the group. Both the agents and the leaders speak of themselves as "jointly" managing the group. In the specific case of our monitoring intervention, some agents for instance mention they choose which farmers to visit together with the leaders. This might stem from the fact that their incentives are quite well aligned: Both see it as their goal to avoid defaults and conflicts in the group. The agent additionally tries to maximize cotton production (which determines the company's profits), and so could be somewhat tougher on input diversion. Nonetheless, both agents and leaders agree that limited amounts of input diversion are acceptable and see it as their goal to prevent excessive moral hazard.[11]

[11] The incentives of the extension agent are thus also quite similar to the incentives of a typical credit officer. A credit officer should care about loan repayment and about

2.4 Results

In this section, we present the impacts of the monitoring intervention. All results are estimated using OLS with standard errors clustered at the village level and controls for unbalanced characteristics.[12]

Field visits and measurements

Table 2.4 presents the impacts of the intervention on the probability that a farmer's field is visited or measured by the agent (columns (1) and (2)), by the leader (columns (3) and (4)) or by either one or the other (columns (5) and (6)). First, in line with the experimental design, agents' visits are more frequent in treatment villages: the share of farmers who received a visit of the extension agent over the campaign increases by 15.9 percentage points. Agents also increased field measurements by 13.6 percentage points. While this was not part of the experimental design, it is not surprising that they took the opportunity of a monitoring visit to measure the farmer's cotton field. Indeed, the cotton company requests that they measure a given number of fields over the campaign.

Second, leaders substantially reduced their visits and field measurements, in almost the same proportion as agents increased monitoring: leaders reduced their visits by 17.3 percentage points and their measurements by 13.1 percentage points. External monitoring thus seems to substitute for internal monitoring by the leaders. In fact, there is no change in the overall probability to receive a field visit or measurement by either the agent or a leader.

While the overall probability to be monitored has not changed, the intervention may have changed the composition of the pool of farmers who are visited. To explore this possibility, in Table 2.5 we compare the characteristics of monitored farmers in control and in treatment villages. Columns (1) and (2) suggest that farmers monitored by the agent are slightly smaller (in terms of area cultivated) and less productive in treatment villages, even though the differences are not statistically significant. In any case, farmers monitored by either the agent or leader have similar

increasing credit sizes when possible. For the agent, avoiding problems in the group is a way to ensure (long term) loan repayment of the group as well as to increase credit area (loan sizes). The difference with a credit officer is that the agent has an interest in increasing expected yields beyond levels at which default becomes impossible.

[12] Given the limited number of clusters, we have also done all these regressions using wild bootstrap with the same level of clustering. This does not affect the results. Additionally, since we sample 13 farmers in each group, the sampling probability of an individual depends on the size of the group. For this reason, we control for the group size in all regressions. (See Solon et al. (2015) for a discussion)

	(1) Agent visit	(2) Agent measurement	(3) Leader Visit	(4) Leader measurement	(5) Agent or leader visit	(6) Agent or leader measurement
Monitored	0.159***	0.136**	-0.173*	-0.131**	0.0327	0.0378
	(0.0566)	(0.0558)	(0.0982)	(0.0635)	(0.0556)	(0.0700)
Constant	0.528***	0.397***	0.276**	0.150*	0.644***	0.436***
	(0.0828)	(0.0755)	(0.125)	(0.0871)	(0.0614)	(0.0931)
N	888	890	796	795	794	795

Standard errors, clustered at the village level, in parentheses. Regressions control for unbalanced characteristics: Whether the HH has a moto, is self-sufficient in cereal production (last 3 years) and is a newly formed HH. There are fewer observations about leader monitoring because this information comes from the leader interview, who were asked only about a random subset of interviewed farmers. Other differences in sample sizes are because of respondents not knowing the answer. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2.4: Effect of the experiment on external and internal monitoring

characteristics in treatment and control villages (columns (5) and (6)) and the frequency of default, risk of default and causes of default are also similar for monitored farmers across treatment and control villages.

Taken together these results indicate that the intervention does not increase the probability of a field visit or measurement (by either a leader or an agent) nor does it affect the type of farmers that are visited. Monitoring is however more likely to be undertaken by the agent rather than by the president and we now investigate how this substitution of internal for more external monitoring changes production and credit outcomes.

Production and default

Table 2.6 presents the impacts of the intervention on farmers' production, default, default risk and type of default in the year of the intervention. As detailed above, the monitoring intervention came as a surprise halfway the agricultural campaign at which point many input decisions were already made (inputs purchased and part of them applied). Actual visits then happened until the end of the campaign at which point almost all decisions were made. The results in Table 2.6 confirm that the intervention has no significant effects on farmer's production behavior. Both cotton yields and the probability that a farmer defaults - meaning that his cotton revenue is smaller than his debt - are remarkably similar in treatment and in control groups (columns (1) and (2)). Likewise, the same proportion of farmers is at risk of default in both groups (a farmer is defined to be at risk of default if a one standard deviation reduction

	(1)	(2)	(3)	(4)	(5)	(6)
		Agent visited		Leader visited		Agent or leader visited
	C	T-C (se)	C	T-C (se)	C	T-C (se)
Characteristics household head and HH						
Age household head	42.90	-0.88 (1.56)	42.89	-0.16 (2.05)	42.67	-0.35 (1.47)
Nr. years of education household head	1.70	-0.12 (0.28)	1.17	0.03 (0.21)	1.47	0.21 (0.22)
Household head has some education	0.29	0.01 (0.05)	0.20	0.04 (0.05)	0.26	0.07 (0.05)
Nr. years household head	16.43	-0.83 (1.20)	18.25	0.34 (1.25)	16.91	-0.44 (0.96)
Nr. years household head is part of credit group	10.73	-0.50 (0.70)	8.82	0.16 (0.69)	9.62	0.03 (0.64)
HH size (nr. people at least 6)	8.70	-0.24 (0.56)	8.81	-1.30** (0.51)	8.09	-0.35 (0.51)
Wealth and food security HH						
HH has television	0.16	-0.01 (0.05)	0.09	0.06 (0.05)	0.10	0.07* (0.04)
HH has bed	0.68	-0.02 (0.14)	0.40	-0.04 (0.16)	0.43	0.00 (0.12)
HH has moto	-0.00	0.00 (0.00)	-0.00	0.00 (0.00)	0.00	0.00 (0.00)
HH has house with solid walls	0.11	0.01 (0.05)	0.11	0.04 (0.08)	0.10	0.04 (0.05)
Consumption proxy (PPI index)	33.63	-0.59 (1.56)	33.25	0.58 (1.71)	32.51	0.64 (1.36)
Self-sufficient in cereal production (last 3 years)	0.00	-0.00 (0.00)	0.00	-0.00 (0.00)	0.00	-0.00 (0.00)
Somebody reduced meal last year	0.30	-0.01 (0.03)	0.39	-0.05 (0.04)	0.37	-0.02 (0.03)
Somebody skipped meal last year	0.18	-0.02 (0.03)	0.24	-0.02 (0.03)	0.23	-0.02 (0.02)
Agricultural production						
Cotton area cultivated 2013	4.87	-0.61 (0.58)	3.16	0.08 (0.57)	3.78	0.08 (0.56)
Average cotton yield (2008-2012)	778.84	-31.75 (39.12)	851.70	-16.59 (39.64)	803.40	-16.82 (35.38)
Cultivated GM cotton 2013	0.75	-0.10 (0.12)	0.63	0.05 (0.11)	0.66	-0.01 (0.11)
Cereal area cultivated 2013	4.70	-0.05 (0.43)	3.44	0.19 (0.37)	3.91	0.27 (0.41)
Credit for cotton production						
Cotton area for credit	4.35	-0.48 (0.65)	3.26	-0.12 (0.65)	3.54	0.10 (0.61)
Total credit for cotton production (in CFA)	436352	-84751 (66171)	340852	-13887 (69825)	354540	-22406 (64381)

Standard errors are clustered at the village level.

Table 2.5: Characteristics of monitored individuals

in his yield would have led to default).[13] Finally, defaults have the same probability to be caused by elements under a farmer's control (columns (4) to (6)). To construct the dependent variables of these last three regressions, we use the leaders' descriptions of the causes of default for all 310 defaults in our sample[14] and call "own fault" defaults that might have been avoided if the farmer had invested more time, attention or inputs to its cotton production. Because this classification is disputable for several causes of default, we construct three measures that are increasingly stringent (70% of defaults are labeled "own fault" with the first definition, 42% with the second and 34% with the third). Table 2.10 in the Appendix presents the frequency of each cause and their classification using the three definitions.[15]

[13]Here, a one standard deviation reduction is based on estimates of farmers' individual yield distribution, assuming farmers have different expected yields but the same variance in yields.

[14]We exclude from these regressions defaults by the leaders since they are reporting about the causes of their own defaults.

[15]In the Appendix we also compare the characteristics of people defaulting (Table 2.13) and of people defaulting because of their own fault (Table 2.15) across control and treatment. They are very similar, again suggesting no change in behavior.

	Yield (kg/ha)	Default	At risk of default	Own fault	Own fault (alt def 1)	Own fault (alt def 2)
Monitored	6.321	-0.0293	-0.0134	0.0314	0.0218	0.0937
	(40.07)	(0.0277)	(0.0459)	(0.102)	(0.0965)	(0.0909)
Constant	736.8***	0.191***	0.415***	0.708***	0.438***	0.351***
	(57.27)	(0.0473)	(0.0704)	(0.0901)	(0.0923)	(0.0731)
N	890	877	877	310	310	310

Standard errors, clustered at the village level, in parentheses. Regressions control for unbalanced characteristics: Whether the HH has a moto, is self-sufficient in cereal production (last 3 years) and is a newly formed HH. The measures own fault are only defined for the 310 defaulting farmers. There are fewer observations for defaults than for yields because some farmers did not know their exact credit amount.

Table 2.6: Effect of external monitoring on yields, defaults and behaviour

These estimates are however not sufficiently precise to rule out small changes in behavior. For instance, the 95% confidence interval of the effect on cotton yields ranges from -74 to 87 kg/ha. With an average cotton yield of 736 kg/ha we can only reject a reduction or increase of more than 10% in yields. We can thus only exclude that the intervention had a large effect on moral hazard in the year of the intervention. Moreover, it may also affect moral hazard in the longer run, but we are not in a position to investigate these effects. We now turn to the impact of the increase of external monitoring on the settling of individual defaults.

Sanctioning of individual defaults

To investigate the consequences of defaults we focus on two subsamples of defaulters: the sample of all farmers whose cotton revenue was smaller than their debt in the control and treatment groups (309 farmers) and the subsample of them who were part of our surveyed sample (112 farmers). Information for the first sample is provided exclusively by the leaders' interviews. The advantage of using the second subsample is that we also know how much credit they obtained the year following the intervention. Clearly, both samples contain only defaulters and are thus selected samples. We discuss the implications of this at the end of this section.

In Panel A of Table 2.7 we see that, following a default, monitored groups less frequently require the farmer to reimburse the debt immediately (as opposed to paying the debt with the next years' cotton revenue): the share of defaulters who repay immediately decreases by about 18 percentage points from 40% in the control group to 22% in the treatment group (column (1)). At the same time, there is a reduction in credit sizes following default. Column (2) reports the change in credit area and column (3) the change in credit size between the year of the intervention and the following year. For defaulters, cotton area financed by credit goes down by 0.5 hectare more in the treatment area. Credit amounts decrease proportionally although the change is non-significant. As detailed above, credit sizes are quite standardized per hectare financed and so credit area and credit amount both capture credit size. In addition, credit area is the more salient and less noisy measure of credit and a significant reduction suggests that there is a reduction in credit for defaulters. Finally, stopping to do cotton is a very drastic change that does not happen frequently. We see an increase in the probability to stop, which is consistent with a reduction in credit, but we do not have enough power to detect reasonably sized effects and will largely disregard the effects on stopping throughout the paper.

An interesting contrast emerges when we investigate the correlation between the type of sanction applied and the causes of defaults (Table 5, Panel B). Results indicate that sanctioning becomes more state-dependent, that is, become more dependent on whether or not the default is related to elements under the farmer's control.[16] Column (1) indicates that within monitored groups, the reduction in reimbursing immediately is concentrated on defaults that are not classified as "own fault". They become 33 percentage points less likely to be sanctioned by immediate reimbursement. This effect is significantly smaller for defaults classified as "own fault". Even if those farmers see a small reduction in reimbursing immediately, this effect is far from significant. Similarly, the reduction in credit seems to be mainly concentrated on people defaulting because of their own fault (significant for credit amount, not significant for credit area). The coefficient on stopping is again estimated too imprecisely to draw conclusions. Tables 2.11 and 2.12 in the appendix confirm that these results generally hold for the three classifications of defaults introduced above.

In short, farmers who are responsible for their default are punished more harshly in treatment villages: they only see a small (non-significant) reduction in reimbursing immediately, but they are significantly more likely to experience a reduction in their loan size (than similar farmers

[16] A possible problem with using a measure like "own fault" is ex-post rationalization, in this case the possibility that the answer about whether or not the default was the farmers' own fault depends on the sanction that has been given. Nonetheless, for this to explain these heterogeneous effects there would need to be differential ex-post rationalization in control and treatment, which seems unlikely.

	(1) Reimbursed Immediately	(2) Credit area (Change)	(3) Credit amount (Change)	(4) Stopped (Year 2)
Panel A: Effect of monitoring on sanctioning				
Monitored	-0.185**	-0.527*	-19461.2	0.0688
	(0.0787)	(0.290)	(31609.9)	(0.0437)
Constant	0.406***	-0.357	-91299.9	0.151*
	(0.106)	(0.395)	(79009.4)	(0.0869)
Panel B: Effect of monitoring on state-dependent sanctioning				
Monitored	-0.339***	-0.0952	89704.4	0.105**
	(0.114)	(0.505)	(58850.7)	(0.0473)
Own fault	-0.135	0.348	134988.3**	0.134**
	(0.130)	(0.336)	(63530.6)	(0.0506)
Monitored X Own fault	0.233*	-0.684	-172299.9***	-0.0643
	(0.136)	(0.598)	(59942.8)	(0.0907)
Constant	0.487***	-0.517	-170010.0	0.0431
	(0.151)	(0.553)	(111489.1)	(0.0848)
N	309	112	108	112

Standard errors, clustered at the village level, in parentheses. Regressions control for unbalanced characteristics: Whether the HH has a moto, is self-sufficient in cereal production (last 3 years) and is a newly formed HH. Reimbursed immediately is defined for all defaulting households, credit measures only for the surveyed ones. Defaults of leaders are excluded from these regressions because these leaders provided themselves the information on the cause of default. Different samples are used for defaulters. The 309 observations come from the leader interview covering all defaults. 112 of these defaulters were surveyed. There are fewer observations on credit amount than on credit area because some respondents did not know their exact credit amount * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2.7: Effect of increased external monitoring on sanctioning

in control villages). In contrast, defaults for elements beyond farmers' control are handled more leniently in treatment villages, leading less often to immediate repayment (while getting the same treatment in terms of credit). The aggregated effect is thus an increase in state-dependency and a shift in the type of sanction used: overall, imposing immediate repayment is less frequent in treatment villages while reducing loan size is more frequent.

The discussion has largely taken all results as causal, but there are several potential identification problems. First, there may be a selection problem since these regressions only involves farmers who defaulted. Defaulters could be different in treatment and control because the intervention could affect moral hazard, for instance because farmers anticipate a change in sanctioning. However, we have argued above that the intervention came largely as a surprise and left little time to change behavior. Additionally, the effects we presented on moral hazard show point estimates, though estimated imprecisely, that are very close to zero (Table 2.6). In Table 2.13 in the Appendix we also compare the characteristics of defaulters in control and treatment and argued that they are very similar. These results suggest that there is not much differential selection between control and treatment and we believe that we do identify a causal effect of the monitoring on sanctioning for the subsample of people who actually defaulted.[17]

The second implicit causal claim concerns state-dependency of sanctioning (farmers are sanctioned differently because they misbehave). Here, there are two potential problems. First, the treatment could have affected the selection of who defaults because of misbehavior (or who is classified as such). Again, moral hazard does not seem to be affected by the treatment and additionally people defaulting because of their own fault are quite similar in control and treatment (see Table 2.15 in the Appendix). The second problem is that, while the monitoring is randomized, "own fault" is not. So, these heterogeneous effects can be generated by any variable that (1) is correlated with "own fault" and (2) causes heterogeneous effects in sanctioning as a result of treatment.[18] While

[17] Since we are interested in sanctioning following default, this is the natural subsample. We cannot rule out however that non-defaulters would have been affected differently, had they defaulted.

[18] Consider this illustrative example. Suppose that educated farmers are more likely both to divert inputs and to challenge the information collected by the group leader in their field (which is used to decide of the sanction). They cannot however challenge the information collected by the external agent, and thus the sanctioning regime would change for them as a result of the intervention. In that case the interaction terms in Table 5 would capture the heterogenous effect of education and not (strictly) that of "own fault". To control for this, we need to control for the interaction between

there are many reasons to believe that those who default for reasons under their control should be systematically different from other defaulters, Table 2.16 in the Appendix suggests that these two groups are very similar. Additionally, when we control for all observable differences, the results remain very similar (Table 2.17 in the appendix). If we are willing to accept that the observed similarity of these different subgroups suggests that they are similar on unobservables, our results suggest a causal effect of the intervention on state-dependency of sanctioning.

Future loans

The results above suggest that the intervention affects credit outcomes for farmers who default on their loan. We now explore whether external monitoring also changes other group members' access to or demand for credit. Several mechanisms may trigger a change in credit demand and supply for individual group members. First, opportunistic behavior might be sanctioned even if it did not lead to default. Second, the fear of future sanctioning of opportunistic behavior may decrease credit demand for farmers who planned to divert inputs. Finally, some farmers may increase their credit demand if they believe that moral-hazard has effectively decreased so that the probability that their revenue will serve to reimburse other members' debt has decreased.

Table 2.8 reports the impacts of the intervention for credit sizes for all surveyed group members. In Panel A we see that there is an overall reduction in credit sizes in treatment villages. Compared to control farmers, treated farmers decrease by an additional 0.27 hectare their cotton area relative to the previous year. The distinction between defaulters and non-defaulters (Panel B) suggests that in monitored villages non-defaulters also decrease their credit area but to a lesser extent than defaulting farmers (non-significant).

As suggested above, both supply and demand factors may explain the decrease in credit sizes. On the one hand, the group (or the leader helped by the extension agent) may now be stricter with farmers suspected to divert inputs away from cotton production. On the other hand, expectations of increased monitoring by the extension agent in the future may discourage input diversion and thereby decrease credit demand. To investigate these two possibilities we use a series of questions that help identify whether farmers faced a binding supply constraint. These questions were asked after the plenary meeting where farmers had introduced their new loan demands. Specifically we ask respondents 1) whether they would

education and the treatment as we do in Table 2.17 in the Appendix.

	(1) Credit area (Change)	(2) Credit amount (Change)	(3) Stopped (Year 2)
Panel A: Effect of monitoring on credit size			
Monitored	-0.275*	-5734.9	0.0144
	(0.162)	(20063.1)	(0.0266)
Constant	0.401	13645.7	0.0905*
	(0.285)	(25320.8)	(0.0514)
Panel B: Effect of monitoring on credit size, by defaulting status			
Monitored	-0.294*	-6968.5	0.0197
	(0.158)	(21104.1)	(0.0234)
Defaulted	-0.408	-115009.4**	0.0817
	(0.334)	(55069.8)	(0.0561)
Monitored X Defaulted	-0.356	-20548.9	-0.0310
	(0.443)	(66263.2)	(0.0726)
Constant	0.442	35471.6	0.0767*
	(0.272)	(25440.7)	(0.0429)
N	872	872	872

Standard errors, clustered at the village level, in parentheses. Regressions control for unbalanced characteristics: Whether the HH has a moto, is self-sufficient in cereal production (last 3 years) and is a newly formed HH. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2.8: Effect of external monitoring on credit sizes

	Defaulters		Entire sample	
	(1)	(2)	(3)	(4)
	Supply constrained?	Agent opposing increase area?	Supply constrained?	Agent opposing increase area?
Monitored	0.230*	0.203	0.0809	0.0753
	(0.129)	(0.172)	(0.0680)	(0.0874)
Constant	0.297*	0.667**	0.417***	0.704***
	(0.157)	(0.245)	(0.0913)	(0.120)
N	119	45	890	290

Standard errors, clustered at the village level, in parentheses. Regressions control for unbalanced characteristics: Whether the HH has a moto, is self-sufficient in cereal production (last 3 years) and is a newly formed HH. Different samples are used for the different regressions. Columns (2) and (4) are restricted to respondents that are supply constrained in Columns (1) and (3), respectively. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2.9: Supply of credit

have liked to increase their credit area by one hectare and - if the answer was negative - 2) whether they would have been allowed to increase their cotton area by one hectare. We classify as supply constrained both those who would have liked to increase their credit area by one hectare and those who could not have increased their area even if they had wanted to (the only difference is that the second group reveals no excess demand for credit). For those who would have liked to increase their credit area by one hectare we additionally asked whether the extension agent had, or would have, opposed their demand.

The results reported in Table 2.9 indicate that the reduction in credit sizes among the people defaulting is driven by a reduction in supply of credit. Furthermore, farmers who were supply constrained are more likely to blame the extension agent for opposing their demand in monitored villages, suggesting that the agent has been more active in the credit meetings.

2.5 Discussion

Two main conclusions emerge from our analysis of the impacts of an increase in external monitoring. First, external monitoring crowds-out internal monitoring by group leaders. Second both types of monitoring are not perfect substitutes when it comes to sanctioning defaults: external monitoring appears to faciliate state-dependent sanctioning in case of default.

In this section, we explore potential mechanisms for these two results in light of the existing literature on the functioning of joint-liability credit groups. We then briefly discuss the welfare implications of these changes.

Crowding out of leader monitoring

The first set of results suggests that the group leaders decrease both field visits and measurements as the extension agent increases his monitoring efforts. Our fieldwork suggests that agents and leaders acquire the same type of information from field visits. If monitoring is costly it may thus not be surprising that leaders would decrease their effort if extension agents increase theirs. Moreover, as we mentioned before, the leaders and agent generally have a good relationship and could to a certain extent coordinate on who needs to be monitored and by whom. However, this result stands at odds with the predictions of the model of joint-liability lending with both peer and external monitoring proposed by Chowdhury (2005), in which external monitoring crowds-in peer monitoring. The theoretical result is driven by strategic complementarity in monitoring efforts: a group member disciplined by monitoring has a greater incentive to monitor his peers because he has more at stake in case they default.

We believe that our experimental design would not allow such strategic complementarities to materialize (if they exist) mainly because the timing of the intervention prevents large responses in terms of investment in cotton production and thus could not starkly reduce moral-hazard. Our results confirm that the intervention has no big effect on productivity or probability of default. Another notable difference with Chowdhury's setting is that we are considering leader monitoring and not strictly peer-monitoring. While peers could theoretically delegate an optimal level of monitoring to the president, in practice incentives may not be perfectly aligned.

External monitoring and internal sanctioning

Our second set of results indicates that there is no perfect substitution between leaders' and extension agents' monitoring efforts: while the overall probability of being monitored has not changed, defaults are handled differently in treatment villages, where the monitoring is more likely to have been undertaken by the extension agents. Specifically, sanctioning appears more state-dependent in treatment villages, whereby farmers are sanctioned more severely when they defaulted for reasons under their control. Moreover, in aggregate, we observe a move away from immediate repayment (less likely in case of "accidental" default in treatment villages) towards greater reductions in loan sizes (more likely in case of "avoidable" defaults in treatment villages).

What can explain this change in sanctioning? We explore two channels : i) a higher level of expertise of the external agent who is more competent at identifying the causes of default and ii) an easier use of the information about the cause of default when it has been collected by an outsider. These explanations are not mutually exclusive. We argue however that both our results and qualitative evidence from the field lend more support to the second explanation.[19]

External monitoring and the type of information collected

Extension agents are agricultural experts who traveled extensively in the cotton area in Burkina and have a deep knowledge of cotton production. One may thus think that their field visits are more informative than a leader's field visit and that they are better at detecting input diversion or predicting yields. If the information on the cause of default is more precise when stemming from the agent, sanctioning of default may become more dependent on this information. However it is not clear why monitoring should then have opposite effects on immediate reimbursement (applied less) versus credit size reduction (applied more).

When we interviewed extension agents, they indicated that they do not get more information from a field visit than the leader would. They believe that the group leaders, who are experienced cotton farmers themselves, are just as competent at detecting signs of input diversion or lack of labour application. In addition, the group leaders have the advantage of knowing the historical record of individual farmers while the rapid rotation of extension agents makes it very difficult for them to acquire this knowledge.

External monitoring and the value of information collected

If the extension agent does not learn more from his visit to a given field than the group leader, the fact that he did the visit (and not the group leader) seems to make a difference. Specifically, it appears that this information is easier to use to settle cases of individual default and

[19] The possibility that the intervention affected the composition of the group of farmers who are monitored (by either the leader or the external agent) was discussed in the result section. While we cannot formally rule this possibility out, we find no indication of such a selection effect.

to sanction when needed, and only when needed. This explanation is supported by the qualitative material obtained during our systematic interviews with group leaders and extension agents.

First, field visits are key to sanction defaults. Group leaders indicated that blaming specific farmers for not investing enough in their cotton production is a serious accusation that needs to be substantiated. To this end, information gathered during a field visit is useful.

Second, using information gathered by agents appears easier than relying on their own field visit. Revealingly, 77% of leaders said that a field visit of a leader is necessary to deal with "moral hazardous" behavior and this number increases to 88% for a visit by the extension agent. When the agent has conducted a field visit and has reported concrete facts, group leaders indicate that they can mention these facts when decisions on sanctioning are taken. In fact, when agents are present during the meeting (which is the case in the group meeting where loan sizes are fixed), they can themselves comment on the production of a given farmers. Our interviews with agents confirmed the importance of their role in the internal settling of defaults. One agent thus mentioned that leaders sometimes even ask him to intervene directly to sanction a member. When doing so he needs to be prudent and motivate the sanction on objective grounds because otherwise "there will be trouble in the village the moment I leave", since farmers will otherwise be well aware that it was the leaders' decision to sanction. A field visit could provide the objective information he needs.

In short our interviews suggest that while the same type of information is gathered by leaders and agents during a field visit, an agent visit has more value when it comes to sanctioning. First it reduces the cost of "pointing fingers" and second it is less disputable because they are largely seen as neutral outsiders. This greater value of the agent's information can account for our results on state-dependent sanctioning and also for the aggregate shift towards less immediate repayment and lower credit size in case of default since the agent is more directly involved with the latter sanction. Indeed the extension agent is present during the meetings where decisions regarding individual credit sizes are taken while he is generally absent of the meeting where debts are settled. In practice, both the agent and the leaders may intervene to oppose a farmers' credit demand. In contrast, the decision regarding immediate repayment is more of an internal affair and, except in exceptional cases where the default leads to serious conflicts in the group, the agent does not intervene. Note also that the meeting where credit demands are introduced and discussed preceeds in time the meeting of debt settlement where decisions regarding debt settlement are taken.

In the credit groups we study here, the leader and all other group members are closely linked by social or kinship ties. They live in the same (small and isolated) village and interact with each other in several other spheres than cotton production. In this context, it is hardly surprising that an outsider (the agent) finds it easier to establish wrongdoing of group members than the leader of the group. In fact, the literature on joint-liability lending acknowledges that social ties have ambiguous impacts on the overcoming of information asymmetries. On the one hand they facilitate information flows and provide an opportunity to use social sanctions. On the other hand they may hinder sanctioning. Wydick (1999) argues that when group members have strong social ties they may be less eager to enforce repayment. Hermes et al. (2005) mention that in joint-liability credit groups, family or friends may be reluctant to use pressure for fear of losing family or friends. Karlan (2007) and Cassar et al. (2007) find that the overall effect of social ties on repayment is positive. In contrast, Ahlin & Townsend (2007) provide some evidence that social ties negatively affect repayment rates. Unfortunately, the precise mechanisms through which social ties hinder or facilitate the functioning of credit groups is not precisely elicited in the existing literature.

Our analysis of credit groups in Burkina Faso suggests that a key obstacle in overcoming ex-ante moral hazard lies in the process of accusing of wrongdoing. The imposition of harsher sanctions is facilitated when the wrongdoing is established by an outsider. Otherwise it seems that all defaults are more likely to be sanctioned in a standardized manner (without distinguishing between accident and wrongdoing).

Welfare implications

What do our results suggest regarding the performance of these joint-liability credit groups? The mix of sanctions changes as a consequence of the monitoring intervention. Immediate repayment becomes less frequent while reduction in credit size is more frequent. Moving away from immediate reimbursement provides useful insurance for defaulters but may be costly in the short-run for other group members as it de-facto delays the full payment of the cotton revenue by at least one year. On the other hand, decreasing the credit size of past-defaulters, especially in the case of opportunistic behavior, may increase the net return from cotton production for the group if these farmers are often at risk of default. The welfare effect of this change in sanctioning regime is thus largely indeterminate.

In contrast, the move towards more state-dependency in sanctioning can be expected to have positive welfare implications. First, the ability to sanction more harshly wrong-doers who could have avoided to default should deter such behavior in the future and should thus decrease the incidence of moral-hazard in the future (provided the intervention would continue). This would decrease the incidence of default and benefit to the group, both as the financial burden associated with individual default lessens and as the tensions generated by the handling of these types of default disappear. Theoretically a decrease in the risk of peer-default may also increase the willingness to make profitable but risky investments, which would further increase profit and income of group members.

Second, the increase in state-dependency in our case also implies that farmers who experience a negative shock and default for reasons beyond their control are less likely to be sanctioned. This implies that risk-sharing across group members de-facto increases. Again this might increase the willingness to make profitable but risky investments. This insurance aspect of jont-liability, and its impact on risk-taking has been studied by Fischer (2013) and Giné et al. (2010) theoretically and using experimental games. These papers indicate that risk-sharing through joint-liability does not necessarily benefit all members (relatively risk-averse borrowers may for example suffer from greater risk-taking by their peers).

Overall the ability to impose state-dependent sanctioning is expected to decrease moral-hazard and increase the insurance value of lending scheme. Quantifying these benefits would require a longer-term intervention to measure its impacts on moral-hazard and risk-taking. This is outside the scope of the current analysis. The costs, on the other hand, can be estimated and are relatively small: Implementing the intervention would require an increase in the interest rate of about 0.23 percentage points[20]. Costs are low because these loans are big (they finance the entire cotton production of the group) and because the intervention is small, requiring only a visit to some group members. It is thus conceivable that the benefits outweigh the costs.

[20]We do not have precise information on the exact costs involved, but we can provide some rough estimates. Agents need to visit each group 8 times. Visiting the farmers (by motorbike) requires 1 litre of gas (750 CFA). The daily wage, including maintenance of the motorbike, is about 5000 CFA and an agent can visit 2 groups per day. The total cost to implement our monitoring intervention for one group is thus about 26 000 CFA (or 40 Euro). By contrast, the average loan size of the groups in our sample is about 11.3 million CFA. The cost of the intervention thus represents an increase of 0.23 percentage points of the interest rate.

2.6 Conclusion

We have argued that a substitution of internal for external monitoring leads to more state-dependent sanctioning. Taking these results at face value, this suggests that external monitoring can be more efficient than internal monitoring. While it is impossible to do a cost-benefit analysis comparing these forms of monitoring, this does suggest that external monitoring in joint-liability groups can be valuable, even when it is more expensive than internal monitoring. This could thus help to understand why it is frequently, and increasingly, used in our context as well as in other credit groups.

We thus make a case for the use of external monitoring in the context of joint-liability loans. A legitimate question is then whether joint-liability with external monitoring has any advantage over individual liability. After all, if costly external monitoring needs to be implemented anyway, it could perhaps also be used to overcome the informational problems in individual liability loans. This is not necessarily the case since joint-liability also makes use of social capital in sanctioning. Joint-liability in combination with external monitoring may thus be needed for group members to be in a position to adequately use these social sanctions to enforce good behavior. This is all the more important as high levels of social capital - which make social sanctions possible - are likely to be positively correlated with high costs of applying punishment.

If our explanation is correct, our results also imply that simply increasing information can not always solve the "informational problem". Even if the group leaders were perfectly informed about the actions of the members, it would not solve moral hazard since they can not act upon this information. A solution to the problem then requires both information and a way to use the information, in this case provided by the actions of the external agent.

Appendix

	Control	sd	Diff: T - C	se	N	Own fault	Own fault (alt def 1)	Own fault (alt def 2)
Did not use enough inputs	0.106	(0.264)	0.102	(0.0753)	310	X	X	X
Did not work enough	0.235	(0.382)	0.0218	(0.0611)	310	X	X	X
Area cultivated too small	0.0101	(0.162)	-0.0299	(0.0207)	310	X	X	X
Bad inputs	0.0114	(0.103)	-0.0106	(0.00709)	310	X	X	
Parasites	0.00548	(0.145)	-0.0163	(0.0187)	310	X	X	
Planted late	0.0128	(0.145)	-0.0149	(0.0201)	310	X	X	
Weeds	0.0548	(0.215)	-0.0468**	(0.0186)	310	X	X	
Bad land	0.0697	(0.177)	0.0483	(0.0347)	310	X		
Not enough labour available	0.194	(0.403)	-0.0181	(0.101)	310	X		
Other	0.00915	(0.177)	-0.00398	(0.0323)	310	X		
Flooding or drought	0.133	(0.425)	0.0215	(0.107)	310			
Lost livestock	0.0900	(0.236)	-0.0122	(0.0422)	310			
Sickness	0.0641	(0.226)	-0.0514***	(0.0182)	310			
Fire	0.00422	(0.0731)	0.0107	(0.0114)	310			

Table 2.10: Comparison of causes of default in treatment and control and definition of own fault variables

	(1) Reimbursed Immediately	(2) Credit area (Change)	(3) Credit amount (Change)	(4) Stopped (Year 2)
Monitored	-0.257**	-0.274	10820.5	0.101**
	(0.102)	(0.369)	(38141.5)	(0.0461)
Own fault (alt def 1)	-0.0656	0.236	145258.0***	0.0870
	(0.0796)	(0.341)	(36763.8)	(0.0638)
Monitored X Own fault (alt def 1)	0.182	-0.596	-78943.4	-0.0835
	(0.129)	(0.570)	(53033.9)	(0.0943)
Constant	0.424***	-0.351	-195948.4*	0.100
	(0.128)	(0.493)	(97386.1)	(0.0834)
N	309	112	108	112

Standard errors, clustered at the village level, in parentheses. Regressions control for unbalanced characteristics: Whether the HH has a moto, is self-sufficient in cereal production (last 3 years) and is a newly formed HH. Reimbursed immediately is defined for all defaulting households, credit measures only for the surveyed ones. Defaults of leaders are excluded from these regressions because these leaders provided themselves the information on the cause of default. Different samples are used for defaulters. The 309 observations come from the leader interview covering all defaults. 112 of these defaulters were surveyed. There are fewer observations on credit amount than on credit area because some respondents did not know their exact credit amount * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2.11: Effect of external monitoring on consequences of default, by own fault (alt def 1)

	(1) Reimbursed Immediately	(2) Credit area (Change)	(3) Credit amount (Change)	(4) Stopped (Year 2)
Monitored	-0.236**	-0.303	-19150.8	0.121*
	(0.0899)	(0.368)	(37788.4)	(0.0644)
Own fault (alt def 2)	0.00579	0.135	85214.1*	0.0676
	(0.0784)	(0.328)	(43061.8)	(0.0673)
Monitored X Own fault (alt def 2)	0.138	-0.580	-33577.9	-0.151
	(0.132)	(0.565)	(56242.7)	(0.116)
Constant	0.396***	-0.255	-145892.1	0.151
	(0.119)	(0.451)	(98853.7)	(0.0980)
N	309	112	108	112

Standard errors, clustered at the village level, in parentheses. Regressions control for unbalanced characteristics: Whether the HH has a moto, is self-sufficient in cereal production (last 3 years) and is a newly formed HH. Reimbursed immediately is defined for all defaulting households, credit measures only for the surveyed ones. Defaults of leaders are excluded from these regressions because these leaders provided themselves the information on the cause of default. Different samples are used for defaulters. The 309 observations come from the leader interview covering all defaults. 112 of these defaulters were surveyed. There are fewer observations on credit amount than on credit area because some respondents did not know their exact credit amount * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2.12: Effect of external monitoring on consequences of default, by own fault (alt def 2)

	Control	sd	Diff: T - C	se	N
Characteristics household head and households					
Age household head	45.00	(12.32)	-3.785**	(1.685)	119
Nr. years of education household head	1.294	(2.275)	0.519	(0.566)	119
Household head has some education	0.193	(0.336)	0.0955	(0.0847)	119
Nr. years household head	15.39	(11.14)	-0.987	(1.712)	119
Nr. years household head is part of credit group	9.255	(6.212)	0.112	(1.391)	118
HH size (nr. people at least 6)	8.090	(4.188)	-0.882	(1.020)	119
Wealth and food security HH					
HH has television	0.0663	(0.458)	0.00829	(0.0908)	119
HH has bed	0.428	(1.044)	0.0892	(0.200)	119
HH has moto	0.484	(0.840)	-0.100	(0.186)	119
HH has house with solid walls	0.0308	(0.177)	0.0931	(0.0626)	119
Consumption proxy (PPI index)	33.27	(11.69)	-0.413	(2.751)	119
Self-sufficient in cereal production (last 3 years)	0.439	(0.504)	-0.0320	(0.0932)	115
Somebody reduced meal last year	0.244	(0.429)	0.0841	(0.0829)	119
Somebody skipped meal last year	0.148	(0.272)	0.0900	(0.0597)	119
Agricultural production					
Cotton area cultivated 2013	2.229	(1.877)	0.480	(0.664)	119
Average cotton yield (2008-2012)	819.3	(255.1)	-20.62	(48.08)	111
Cultivated GM cotton 2013	0.602	(0.502)	-0.0497	(0.118)	119
Cereal area cultivated 2013	3.312	(2.279)	0.408	(0.650)	119
Credit for cotton production					
Cotton area for credit	2.479	(2.955)	0.105	(0.830)	119
Total credit for cotton production (in CFA)	284730.1	(377866.7)	-38674.6	(102273.4)	114
Monitoring by leader and agent					
Agent visited field	0.232	(0.447)	0.279***	(0.0940)	119
Agent measured field	0.172	(0.368)	0.200**	(0.0914)	119
Leader visited field	0.317	(0.502)	-0.129	(0.180)	105
Leader measured field	0.0283	(0.269)	0.120	(0.0969)	105
Leader or agent visited field	0.590	(0.486)	0.0353	(0.164)	105
Leader or agent measured field	0.271	(0.437)	0.169	(0.112)	105

See the note below Table 2.1 for an explanation of the missing variables. Standard errors are clustered at the village level.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2.13: Comparison of defaulters in treatment and control.

	Non-defaulters	sd	Diff: Defaulters - Non-defaulters	se	N
Characteristics household head and households					
Age household head	44.03	(13.01)	-0.518	(0.980)	890
Nr. years of education household head	1.128	(2.488)	-0.264	(0.309)	890
Household head has some education	0.241	(0.434)	-0.0854*	(0.0432)	890
Nr. years household head	16.26	(11.70)	-0.266	(1.116)	889
Nr. years household head is part of credit group	9.882	(6.354)	0.0293	(0.685)	888
HH size (nr. people at least 6)	8.836	(5.350)	-1.156**	(0.476)	890
Wealth and food security HH					
HH has television	0.229	(0.529)	-0.0709	(0.0478)	890
HH has bed	0.635	(1.338)	-0.153*	(0.0888)	890
HH has moto	0.813	(1.068)	-0.339***	(0.0897)	890
HH has house with solid walls	0.131	(0.381)	-0.108***	(0.0290)	890
Consumption proxy (PPI index)	37.66	(12.55)	-4.232***	(1.117)	890
Self-sufficient in cereal production (last 3 years)	0.607	(0.483)	-0.139***	(0.0456)	852
Somebody reduced meal last year	0.186	(0.392)	0.0874**	(0.0393)	890
Somebody skipped meal last year	0.110	(0.298)	0.0210	(0.0311)	890
Agricultural production					
Cotton area cultivated 2013	3.850	(3.506)	-1.310***	(0.277)	890
Average cotton yield (2008-2012)	872.8	(257.9)	-80.12***	(24.19)	851
Cultivated GM cotton 2013	0.737	(0.495)	-0.112**	(0.0542)	890
Cereal area cultivated 2013	4.319	(3.256)	-0.660**	(0.293)	887
Credit for cotton production					
Cotton area for credit	3.922	(3.780)	-1.156***	(0.331)	890
Total credit for cotton production (in CFA)	399672.2	(404052.3)	-100822.8**	(37458.7)	877
Monitoring by leader and agent					
Agent visited field	0.505	(0.500)	-0.128**	(0.0527)	888
Agent measured field	0.363	(0.485)	-0.129**	(0.0499)	890
Leader visited field	0.308	(0.495)	-0.0546	(0.0726)	796
Leader measured field	0.116	(0.365)	-0.0221	(0.0545)	795
Leader or agent visited field	0.642	(0.446)	-0.0822	(0.0739)	794
Leader or agent measured field	0.427	(0.498)	-0.128**	(0.0588)	795

See the note below Table 2.1 for an explanation of the missing variables. Standard errors are clustered at the village level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2.14: Comparison of defaulters and non-defaulters (in entire sample).

	Control	sd	Diff: T - C	se	N
Characteristics household head and households					
Age household head	47.43	(12.99)	-5.595**	(2.522)	75
Nr. years of education household head	0.796	(1.767)	1.008	(0.604)	75
Household head has some education	0.105	(0.273)	0.191*	(0.0972)	75
Nr. years household head	16.77	(12.10)	-1.811	(2.460)	75
Nr. years household head is part of credit group	10.25	(6.189)	-1.162	(1.760)	74
HH size (nr. people at least 6)	8.659	(4.759)	-1.117	(1.200)	75
Wealth and food security HH					
HH has television	0.110	(0.563)	-0.0207	(0.117)	75
HH has bed	0.368	(0.847)	0.198	(0.234)	75
HH has moto	0.614	(0.887)	-0.227	(0.215)	75
HH has house with solid walls	0.00577	(0.162)	0.0551	(0.0511)	75
Consumption proxy (PPI index)	32.94	(11.72)	-1.277	(3.230)	75
Self-sufficient in cereal production (last 3 years)	0.357	(0.500)	-0.0907	(0.117)	73
Somebody reduced meal last year	0.256	(0.446)	0.0883	(0.112)	75
Somebody skipped meal last year	0.179	(0.273)	0.108*	(0.0622)	75
Agricultural production					
Cotton area cultivated 2013	2.296	(2.063)	0.175	(0.780)	75
Average cotton yield (2008-2012)	810.9	(234.0)	-2.943	(42.46)	71
Cultivated GM cotton 2013	0.623	(0.500)	-0.0462	(0.144)	75
Cereal area cultivated 2013	3.607	(2.530)	-0.0804	(0.722)	75
Credit for cotton production					
Cotton area for credit	2.751	(3.538)	-0.536	(1.069)	75
Total credit for cotton production (in CFA)	263957.7	(429868.0)	-80647.3	(124020.0)	73

See the note below Table 2.1 for an explanation of the missing variables. Standard errors are clustered at the village level.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2.15: Comparison of people defaulting because of their own fault in treatment and control.

	Default not own fault	sd	Diff: Own fault - Not own fault	se	N
Characteristics household head and households					
Age household head	44.44	(10.61)	-1.465	(2.173)	119
Nr. years of education household head	1.556	(2.715)	-0.0631	(0.430)	119
Household head has some education	0.230	(0.370)	0.00373	(0.0749)	119
Nr. years household head	14.66	(8.986)	0.450	(1.916)	119
Nr. years household head is part of credit group	9.372	(6.560)	-0.0994	(1.183)	118
HH size (nr. people at least 6)	7.696	(4.123)	0.0356	(0.751)	119
Wealth and food security HH					
HH has television	-0.0119	(0.255)	0.117	(0.0697)	119
HH has bed	0.553	(1.087)	-0.126	(0.164)	119
HH has moto	0.350	(0.685)	0.132	(0.130)	119
HH has house with solid walls	0.112	(0.321)	-0.0600	(0.0643)	119
Consumption proxy (PPI index)	31.12	(11.54)	2.834	(2.778)	119
Self-sufficient in cereal production (last 3 years)	0.669	(0.457)	-0.336***	(0.0948)	115
Somebody reduced meal last year	0.223	(0.424)	0.0800	(0.0837)	119
Somebody skipped meal last year	0.164	(0.291)	0.0316	(0.0510)	119
Agricultural production					
Cotton area cultivated 2013	2.625	(2.882)	-0.279	(0.540)	119
Average cotton yield (2008-2012)	807.9	(264.7)	2.951	(51.24)	111
Cultivated GM cotton 2013	0.654	(0.505)	-0.103	(0.0947)	119
Cereal area cultivated 2013	3.619	(3.172)	-0.195	(0.562)	119
Credit for cotton production					
Cotton area for credit	2.563	(2.644)	-0.0566	(0.695)	119
Total credit for cotton production (in CFA)	306941.9	(314010.6)	-54301.3	(82017.0)	114

See the note below Table 2.1 for an explanation of the missing variables. Standard errors are clustered at the village level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2.16: Table comparing people who default because of their own fault and those who do not

	(1) Reimbursed Immediately	(1') Reimbursed Immediately	(1b) Reimbursed Immediately	(2) Credit area (Change)	(2b) Credit area (Change)	(3) Credit amount (Change)	(3b) Credit amount (Change)	(4) Stopped (Year 2)	(4b) Stopped (Year 2)
Monitored	-0.339***	-0.330***	-0.250	-0.0952	0.580	89704.4	27811.0	0.105**	0.181
	(0.114)	(0.0857)	(0.213)	(0.505)	(1.018)	(58850.7)	(131213.6)	(0.0473)	(0.254)
Own fault	-0.135	-0.0490	-0.0368	0.348	0.378	134988.3**	101373.4	0.134**	0.105**
	(0.130)	(0.0867)	(0.103)	(0.336)	(0.387)	(63530.6)	(63987.9)	(0.0506)	(0.0446)
Monitored X Own fault	0.233*	0.106	0.0968	-0.684	-0.940	-172299.9***	-166055.6**	-0.0643	-0.0125
	(0.136)	(0.103)	(0.110)	(0.598)	(0.713)	(59942.8)	(79041.3)	(0.0907)	(0.102)
Constant	0.487***	0.190	-0.118	-0.517	1.458	-170010.0	46656.6	0.0431	-0.313
	(0.151)	(0.156)	(0.266)	(0.553)	(1.032)	(111489.1)	(209554.2)	(0.0848)	(0.191)
Additional controls	No	No	Yes	No	Yes	No	Yes	No	Yes
Surveyed farmers only	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	309	112	112	112	112	108	108	112	112

Standard errors, clustered at the village level, in parentheses. All regressions control for unbalanced characteristics: Whether the HH has a moto, is self-sufficient in cereal production (last 3 years) and is a newly formed HH. Columns (1), (2), (3) and (4) replicate the same columns in Table 2.7 Panel B. Before adding controls, column (1') restricts the sample to surveyed farmers, for whom controls are available. The additional controls in the other columns are: Age household head, Household head has some education, Somebody skipped meal last year, Monitored X self-sufficient in cereal production, Monitored X newly formed HH, Monitored X Household head has some education, Monitored X Household head close family of a village leader, Monitored X Consumption proxy, Monitored X Cotton yield 2013, Monitored X Cotton area cultivated 2013 and all base levels for the interacted controls. The non-interacted controls are the significant differences between people misbehaving reported in Table 2.15. The interacted controls are variables that could cause the heterogeneous effects in monitoring. We include all variables that correlate significantly with "own fault" (see Table 2.16), as well as some that do not correlate significantly but could plausibly cause the heterogeous effects. Results are virtually unchanged by controlling. The heterogeneous effect on reimbursing immediately loses significance because of the sample restriction, but is virtually unaffected by the controls. Globally, the point estimates are similar with controls and, if anything, move towards somewhat more state-dependency in sanctioning. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2.17: The effect of external monitoring on state-dependent sanctioning, with and without controls.

Chapter 3

The demand for micro-insurance: A literature review[1]

[1] This chapter is co-authored with Ombeline De Bock and Jean-Philippe Platteau.

Abstract

Micro-insurance has recently received much attention as a promising tool to protect poor individuals from important shocks. Yet, voluntary demand from people has been low, shedding doubt on the viability of micro-insurance as a useful risk-management tool. To better understand this puzzle, this paper reviews both the theoretical and empirical literature on the demand for insurance. While people's lack of understanding of insurance does seem to limit the demand for it, several more fundamental factors, such as price, quality, limited trust in the insurer, and liquidity constraints also seem to have an important role in explaining the puzzle.

3.1 Introduction

Micro-insurance - or the insurance for the poor - has been considered as "the next revolution" in addressing risks and vulnerability in low-income countries (Morduch, 2006). In particular, huge investments have been made in this revolutionary tool in the last decade by several development agencies (among which USAID and the Gates foundation) in the hope of breaking the circle of poverty and offering a reliable protection to the poor.

Its name echoes the well-known micro-credit phenomenon, on purpose. Both concepts have in common a specific attention to low-income households in the developing world. They, moreover, try to solve a market imperfection which is identified as a major cause perpetuating poverty. However, micro-insurance is an even more complex concept than micro-credit. First, it implies paying a regular premium in return for an uncertain payout. Second, it is mostly conceived as a set of individual, rather than group-based, contracts where some enrolees benefit from a compensation while others do not. Finally, micro-insurance is far from being homogeneous: It concerns a wide variety of risks and takes a lot of different forms.

The focus of this review is on low-income countries, where adverse shocks are frequent, and risk-pooling mechanisms and self-insurance strategies are imperfect. As poor individuals also display a relatively high level of risk aversion, the demand for micro-insurance products is thus expected to be high. However, the evidence is disappointing: subscription to the widely subsidized insurance schemes is low, rarely above 30%. And while this could still be seen as a reasonable rate for a new product, renewal rates are also low, ranging from 10 to 70%. At such rates, insurances cannot be sustainable and will fail to deliver the benefits it promises. The question of this paper is thus: Why is demand and renewal for micro-insurance so poor?

The present paper addresses this puzzle both from a theoretical point of view and by reviewing the empirical evidence on the factors influencing demand for insurance. Given the numerous number of studies published on micro-insurance in the past ten years, and the diverging results obtained, we believe this review is not only necessary, but also timely. We will focus both on demand and renewal of micro-insurance. Renewal is a topic that has received far less attention than demand despite its importance in promoting a sustainable insurance scheme. While demand and renewal are, of course, related, the decisions to purchase an insurance with or without having experienced it are not the same and we believe that explaining low renewal rates is critical for a correct assessment of the potential of micro-insurance.

Our literature review is extensive, covering a wide range of topics including behavioural models, supply deficiencies of insurances and the role of existing substitutes of insurances[2]. It considers qualitative as well as quantitative, and theoretical as well as empirical papers that study the demand for insurance in developing countries. However, given the extensive recent literature on this subject, we only present the papers which bring the most robust evidence on the different aspects of demand. For this reason, we do not rely much on studies investigating the hypothetical willingness-to-pay (WTP) since the methodology used to elicit hypothetical demand presents great challenges. In particular, these studies systematically produce overly optimistic estimates of the demand for insurance and, in some cases, the WTP even fails to correlate with actual demand (McIntosh et al., 2013).

Various types of insurance are considered, among which two main categories may be distinguished: contracts insuring the subscriber against the risk of incurring medical expenses and insurance contracts against harvest losses. Among crop insurance schemes, those based on an index have recently received considerable attention. The fundamental difference with classical insurance lies in the nature of the event that triggers the payment of compensation. The index insures against the occurrence of an easily identified event that correlates with an expected decrease in the incomes of the farmers in a given area. In weather index insurances, for instance, it is the level of rainfall, rather than the observed damage, that triggers the payout.

Although the evidence is far from decisive, several lessons can be drawn from this review. Understanding the concept of insurance is not an easy task for individuals but there exists a wide range of alternative explanations as to why demand for conventional insurance schemes is so low. A lack of trust in the institution delivering the insurance, or in the specifics of the product may significantly decrease uptake. Similarly, the frequency of payouts, the quality of the product and liquidity constraints are pointed to as important factors affecting demand.

A major feature of the present review is the attention devoted to

[2]For other reviews on the topic of micro-insurance, see Miranda & Farrin (2012) andCarter et al. (2014) on index-based insurances ; Ekman (2004) on community-based health insurances ; De Bock & Ontiveros (2013) on the impact of micro-insurance ; and Eling et al. (2014) on the demand for micro-insurance, with a particular focus on linking it to demand for insurance in developed countries. The main difference of this paper with the latter one is a focus on economic theory and the wide scope of the paper, extensively discussing topics such as behavioural models and substitutes for insurance.

economic theory in order to shed light on empirical findings. Therefore, we start by presenting various theories, including behavioural ones, which help to explain observed insurance demand. In the same section, we also present the evidence on other characteristics of subscribers, such as their understanding of insurance and risk aversion, which affect demand. Next to demanders' characteristics, supply deficiencies such as high prices or basis risk matter. These are treated in Section 3. We then discuss how a lack of trust in insurance affects uptake. Finally, in Section 5 we consider different substitutes for insurance such as informal risk-sharing and credit. Section 6 concludes.

3.2 Low demand arising from characteristics of potential subscribers

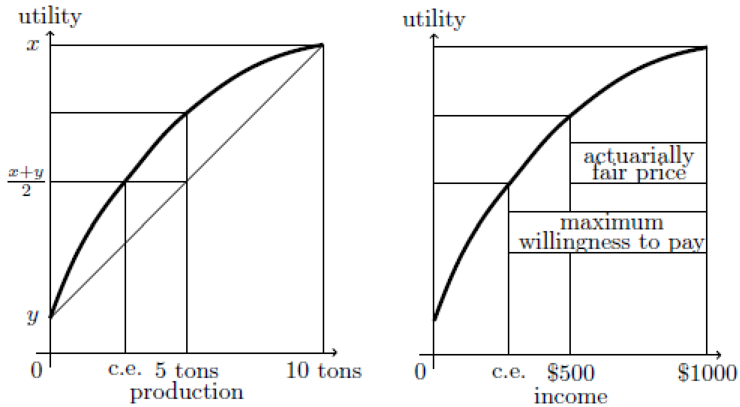
3.2.1 Behavioural explanations

Before assessing which factors affect the demand for micro-insurance, it is useful to discuss the available theories. The dominant one is the expected utility theory, which is the standard theory of decision making under uncertainty. However, in an attempt to explain the low demand for insurances, researchers have increasingly resorted to behavioural models. In this section, we first summarize the central point of the expected utility theory, and then turn to the most important alternative setups based on various behavioural assumptions.

To see why insurance can be valuable under expected utility theory, we need to take a closer look at the level of satisfaction or utility each individual derives from subscribing to the policy. According to this theory, people try to maximize their expected utility when deciding whether to purchase the insurance product. If marginal utility is decreasing, that is, if greater consumption leads to more utility but that the increase in utility is smaller for each additional increase in consumption, the utility function of the agent is concave. This specific feature of the utility function gives rise to risk aversion. Indeed, as can be seen in the utility function in Figure 3.1(a) (Patt et al., 2009), a risk averse farmer will always prefer receiving a definite amount (the average of two possible harvests) over a risky situation in which each possible harvest is equally likely (in which case his utility is the average of the two levels of utility that he could experience).

The readiness to pay for receiving a certain amount allows an insurance market to emerge. Indeed, because of the aversion for uncertain outcomes, risk averse individuals will be willing to pay more than the

Figure 3.1: Utility, risk-aversion and willingness to pay for insurance



(a) Utility function (b) Willingness to pay

actuarially fair price of the insurance - the price which is equal to the average amount the insurance will pay out - in order to receive a compensation in case the harvest fails. A key concept is that of certainty equivalent (c.e.) which measures the certain income that the individual would consider equivalent (from a utility standpoint) to the lottery he (she) wants to avoid (for example, 3 tons instead of 50 percent chance of obtaining 0 tons and 50 percent chance of obtaining 10 tons, that is, a lottery of which the average outcome is equal to 5 tons).

Figure 3.1(b), an example elaborated by Patt et al. (2009), depicts the case in which each ton of production is worth USD 100. Without any production shocks, the farmer may expect to harvest 10 tons of cereals and get 1000 USD. However, in case of a drought, the farmer loses his entire production and earns USD 0. When there is a 50% probability of drought, the average loss is USD 500. If an insurance pays USD 1000 in case of harvest failure, the figure shows that a risk averse individual will be willing to pay up to the difference between USD 1000 and his certainty equivalent for the insurance, which is more than the actuarially fair price of the insurance.

From the tight framework of rational expected utility theory, a central result is easily derived: the more risk-averse an individual, the higher the risk premium he (she) is willing to pay to get insured against shocks. It is thus readily checked from the figures that the more concave the utility function (i.e., the stronger the aversion toward risk) the larger the distance between the distance between the certainty equivalent level and the maximum income.

We can now consider alternative theories based upon behavioural assumptions that differ from the conventional rationality assumption underlying the canonical model described above. We focus on those that are directly relevant to explain low demand for insurance among poor people.

Ambiguity aversion

A first alternative approach is ambiguity aversion theory. Many people appear to be ambiguity averse, that is, dislike being uncertain about the likelihood with which events will occur (Ellsberg, 1961)[3] . This is different from typical risk aversion: Instead of disliking that outcomes are uncertain, an ambiguity averse individual dislikes being uncertain about the distribution of outcomes.

Both Elabed & Carter (2015) and Bryan (2013) argue that ambiguity aversion might limit take-up of insurance. While people know what to expect when not buying insurance, the choice to purchase insurance comes with plenty of ambiguity, for example about the exact trustworthiness of the insurer or the exact coverage of the contract. An ambiguity averse individual would then evaluate the insurance contract by assuming the least conceivable trustworthiness and coverage, and conclude that insurance is not very valuable.

Elabed & Carter (2015) argue that ambiguity aversion has the effect of discouraging demand for index insurances in particular, because this type of insurance suffers from basis risk. Basis risk is the risk that an insurance does not pay out, even though there are losses. For instance, an index insurance based on rainfall will not pay out if there is a pest problem. In this case, there are two levels of uncertainty - about having losses, but also about receiving a payout in case of losses – and this is particularly unappealing for an ambiguity averse individual. Since almost 60% of the potential micro-insurance clients they have interviewed revealed themselves to be ambiguity averse, they argue that high levels of basis risk can substantially reduce demand.

[3] Ambiguity aversion is still best explained by the original experiment of Ellsberg (1961): An individual faces two urns. The first contains 10 balls, 5 red and 5 blue; the second also contains 10 red or blue balls, but in unknown proportions. The individual can choose a color and an urn to draw a ball from, and wins if he draws the chosen color. An individual following subjective expected utility theory should believe that in the second urn, either for red or for blue, he has at least 50 percent probability of drawing this color. Nonetheless, most individuals strictly prefer drawing from the first urn because the odds are not ambiguous; they are ambiguity averse.

Hyperbolic discounting

Every potential client has an intrinsic discount rate, that is, a degree of preference for present consumption. Paying a premium today and only receiving a payout in the future thus implies an opportunity cost of not having used the money during the period in between. Thus, a risk-neutral agent with a positive discount rate has no incentive to take-up an actuarially fair insurance. When people exhibit hyperbolic discounting, or time-inconsistent preferences, this effect is even stronger. On top of preferring consumption sooner rather than later, hyperbolic discounting implies a preference to consume today, simply because it is today. That is, one might systematically prefer to receive 2 dollars in a week and one day instead of 1 dollar in a week and at the same time prefer to receive 1 dollar today over 2 dollars tomorrow. Hyperbolic discounting can lead to a low demand for insurance because the premium needs to be paid today, but the potential benefits are experienced only in the future. It could thus be seen as a lack of self-control. Indeed, even when an individual with time-inconsistent preferences is willing to purchase insurance - and would commit, if possible, to do so in the future - he might well decide not to do so at the moment the payment needs to be made.

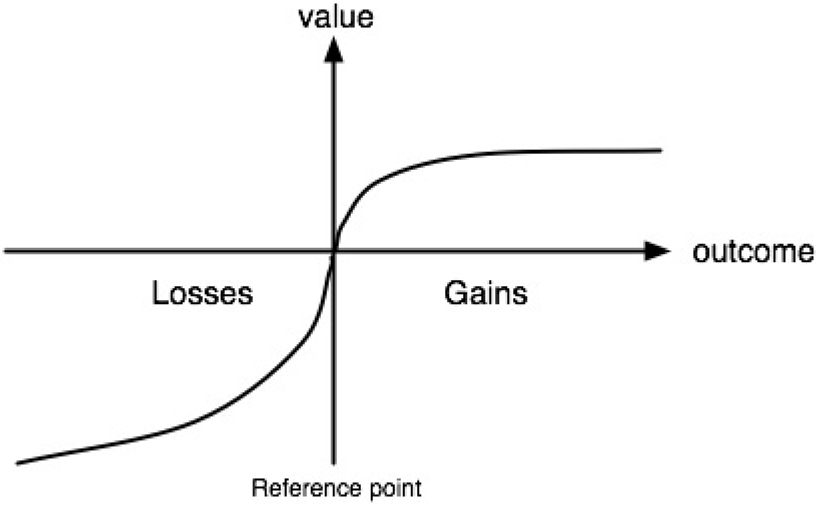
Nonetheless, Ito & Kono (2010) estimate the demand for insurance for a small group of people exhibiting hyperbolic preferences and find that they are more likely to purchase health insurance. They argue that these people use the insurance as a commitment device: having time-inconsistent preferences, they know they will have difficulties to save for uncertain health expenditures. The insurance, which can be seen as a prepayment of health expenditures, is thus especially valuable.

Loss-aversion and prospect theory

Kahneman & Tversky (1979) argue that people exhibit loss-aversion, that is, they experience more disutility for a loss, than they experience utility for a gain of the same amount. Thus, unlike in expected utility theory, the framing of a change in wealth matters: feeling that a loss of a certain amount has been avoided gives more utility than simply gaining the same amount.

Such loss-aversion can influence insurance behaviour. Marketing insurance as preventing a loss ("don't lose your property, buy insurance to be covered in case of emergencies"), rather than allowing a gain ("increase your peace of mind, buy insurance to be covered in case of emergencies"), could increase people's perception of the value of insurance (Ganzach & Karsahi, 1995).

Figure 3.2: Utility function under prospect theory



Loss aversion is, in fact, only one part of a more elaborate theory of decision making under uncertainty, known as prospect theory (Kahneman & Tversky, 1979; Tversky & Kahneman, 1992). On top of the assumption of loss aversion, prospect theory is based on two additional assumptions. First, the utility function for losses is convex, implying risk-seeking, while the one for gains is concave (see Figure 3.2). Second, the probability distribution with which expected utility is evaluated is not the true one but is weighted such that the worst (low probability) outcomes are overweighed and other outcomes are underweighted.

Together, these assumptions imply that in the domain of losses - which is the relevant domain in the context of insurance - individuals are risk-averse for the worst outcomes and risk-seeking otherwise. It is this risk-aversion for the worst events which makes insurance valuable under prospect theory. The reason for this is thus quite different than under expected utility theory. Instead of the concavity of the utility function, it is the overweighting of the worst events which makes insurance valuable (Wakker et al., 1997).

If overweighing of the worst outcomes is indeed an important reason for purchasing insurance, this could explain why certain insurances are in low demand. Perceptions of what is covered by the insurance then play a decisive role in the take-up. In particular, when an insurance cannot guarantee full coverage, some of the worst events are not covered. As these events are overweighed most, this lack of coverage gives a disproportional amount of disutility and the insurance is perceived as less valuable. In a WTP study in the USA, Wakker et al. (1997) show that people demand more than a 20% reduction in price when an insurance has a 1% probability of non-payout. The implication is that insurance is valued more if, instead of simply reducing a risk, it gives the impression of removing it completely.

The evidence for prospect theory is mixed. Ito & Kono (2010) assess the attitudes of individuals vis-a-vis risk and find that a large number of them are risk-loving in the domain of losses. These loss risk-lovers also seem to be somewhat less likely to purchase insurance. On the other hand, Dercon et al. (2011) find that people are risk-averse in the domain of losses, and actually even more than in the domain of gains. As prospect theory assumes that people are risk-averse for big losses and risk-seeking for smaller ones, it is however debatable whether these results support or contradict prospect theory. Concerning the weighting of probabilities under prospect theory, Clarke & Kalani (2011) actually find that insurance take-up decisions in a game are better explained by the underweighting of extreme events than by the overweighting assumed by prospect theory.

Discontinuity of preferences at certainty

We argued above that a strong dislike of an insurance that has a small probability of non-payout can be explained by prospect theory. However, another explanation is provided by the theory of discontinuity of preferences at certainty, which posits that people dislike uncertainty beyond what is predicted by expected utility theory (Andreoni et al., 2010; 2012). That is, when comparing an insurance with either 0%, 1% or 2% probability of non-payout, an individual with discontinuous preferences might strongly prefer the insurance with 0% non-payout, and be quite indifferent between 1% and 2% non-payout. This is impossible for an expected utility maximizer: being almost indifferent between 1% and 2% non-payout implies being almost indifferent between 0% and 1% as well. An individual with discontinuous preferences can thus have a special preference for certain outcomes.

Serfilippi et al. (2015) show that such discontinuous preferences at certainty can matter for insurance demand. In an insurance game with farmers in Burkina Faso they frame the insurance either with or without the certainty of having to pay the premium: in the uncertain frame, the premium does not need to be paid when there is a payout. Individuals who reveal themselves to have discontinuous preferences (29% of the sample) are willing to pay substantially more when the premium payment is framed as uncertain, while others are not. These individuals therefore seem to dislike the certainty of paying the premium.

Demand could thus be potentially increased by stating that the premium only needs to be paid when there is no payout, but this is only rarely possible. However, discontinuity of preferences can also affect demand in other ways. Most importantly, the insurance should give the impression that it is sure to yield a payout when needed. An insurance that is not trusted or that has high levels of basis risk could be in low demand among people with discontinuous preferences.

3.2.2 Education and the understanding of insurance

Understanding of insurance

A huge challenge for micro-insurance schemes is that they must be properly understood by potential and actual clients. The core concept of insurance - spending money in return for an uncertain payout covering a hypothetical event – is, indeed, not straightforward. This is perhaps best illustrated by the demand of many newly insured to receive their premium back in case no payout occurs (Platteau, 1997; Basaza et al., 2008). Even in the United States, it has been argued that consumers of insurance treat insurance as a short-term investment so that, if they have not collected on their policy over several years, they feel that the premiums paid have been wasted (Kunreuther et al., 2013, pp. 104-5, 117-118). Such misunderstandings can threaten the sustainability of the insurance: Platteau & Ontiveros (2013) find that a lack of payouts leads people to drop out of the insurance, except if they have a good understanding of insurance.

Platteau (1997) gives a plausible explanation for the lack of understanding of the insurance concept. He argues that in traditional risk-sharing arrangements members of traditional rural communities "are guided by a principle of balanced reciprocity (they expect a return from any contribution or payment they make) rather than by a true logic of mutual insurance. More precisely, they do not conceive of insurance as a game where there are winners and losers and where income is redistributed between lucky and unlucky individuals." If micro-insurance is evaluated with such a logic of balanced reciprocity, the demand to be reimbursed when there is no payout appears fair rather than extravagant.

However, even if people evaluate insurance in a framework of balanced reciprocity, this does not make it impossible for an insurance scheme to be successful. Most importantly, an insurance needs to pay out often enough to ensure a feeling of some reciprocity, which does not necessarily need to be perfectly balanced. By mixing different risks, and thus covering many of them, an insurance could achieve such frequent payouts. In another way, a life insurance, by guaranteeing a payout at some point, guarantees reciprocity. However, other types of insurance, such as agricultural insurance, often rely on the principles of spreading (big) risks among many clients and redistributing money from lucky to unlucky individuals through infrequent payouts. For these, one needs to accept the logic of insurance to purchase it over extended periods of time.

Beyond the basic concept of insurance, other features of insurance operations, such as the pooling of contributions, the presence of a deductible, or the exact perils covered by the contract can also raise mixed understanding among the beneficiaries. From this perspective, index-based insurance poses most problems. First, one needs to understand the basic principles of insurance. Second, one can easily be deceived by the absence of a payout in case of a loss because of basis risk. Indeed, to fully understand basis risk one typically needs a grasp of concepts such as the average and have an idea of the amount of correlation between the index and individual outcomes, which is not evident. Finally, the index itself can pose serious problems. This may happen when, for instance, it is expressed in millimetres or computed based on satellite imagery.

The logic of insurance does not therefore seem easy to internalize. In the following, we discuss various aspects on which to focus in order to improve the understanding of insurance.

Education

Various empirical studies conclude that the more years of schooling respondents have completed, the more likely they are to enrol in the insurance scheme (Giesbert et al., 2011; Giné et al., 2008; Jehu-Appiah et al., 2012; Jowett, 2003; Schneider & Diop, 2004). This is consistent with the idea that better educated people may have a better understanding of the insurance product and are therefore more likely to purchase it. More specifically, Giesbert et al. (2011) note that education might stimulate demand by increasing financial literacy, and that the effect of education on demand may thus vanish once financial literacy is controlled for. When controlling for financial literacy, Giné et al. (2008) indeed do not find a significant impact of years of school attendance on take-up. Platteau & Ontiveros (2013), on the other hand, when controlling for the level of understanding of insurance, find a non-monotonous effect of schooling on renewal: It is negative in the first years and becomes positive only when a sufficient level of education is reached. They argue that this is caused by the fact that more educated people tend to be more frustrated when confronted with a dysfunctional insurance, whereas at relatively high levels of education reduced transaction costs for renewing the contract for literate people may outweigh this effect. Education could thus affect demand and renewal in different ways.

For this reason, education should not be simply seen as a proxy for understanding of insurance or financial literacy. Thus, Platteau & Ontiveros (2013) find almost no correlation between education and the level of understanding of insurance. Similarly, Gaurav et al. (2011) find no correlation between education and financial literacy. They moreover bring evidence that low financial awareness of the respondents lowers the probability of adopting the rainfall insurance. Interestingly, Cole et al. (2013) isolate the understanding of probabilities when testing the respondents' financial awareness. The authors observe that financial literacy per se has no effect on insurance demand, but being comfortable with the probabilistic concept seems to be strongly correlated with the decision to purchase insurance.

Improving knowledge through training in financial literacy

Given the apparent problems people meet in getting a proper grasp of the concept of insurance, and the often positive correlation between financial literacy and demand, financial literacy trainings are widely expected to substantially increase demand. However, many studies have now tested the impact of different types of trainings, and the evidence is mixed. While financial literacy training for complex index-based insurances generally seems to have some positive impact on demand, studies for health insurance have not found any significant impact. As financial literacy training programs vary substantially in terms of approach and intensity, it is worth discussing the different projects in more detail.

While training in insurance literacy can increase the knowledge of the product specificities, its impact on the purchase of insurance is unclear. For instance, Gaurav et al. (2011) find a positive effect of general training in financial literacy but only for people with low initial levels of financial literacy. Moreover, carrying out a demonstration of how the millimetre threshold triggers a payout of their weather index-insurance had no effect. As another example of the complexity and diversity of factors to take into account when assessing the effectiveness of literacy training, a study by Giné et al. (2014), who distributed comics as a way to provide financial literacy training, find a substantial effect on demand only when enough people in the village received the comics.

As mentioned before, the studies on financial literacy training for health insurance find no effect. Neither a three-hour training session about the insurance and general financial management (Bonan et al., 2012), nor several educational modules varying in intensity (Schultz et al., 2013) appear to have had an impact on the demand for insurance. Adopting another approach, Dercon et al. (2011) randomly assign people to so called "study circles" which consist in regular meetings of a study group in which written materials about insurance are discussed. This kind of participative training session too did not have any significant impact on the purchase of health insurance either.

Finally, the perception of participants about different kinds of training seems to differ. Patt et al. (2010) argue that playing games does not improve understanding more than what a traditional training would achieve, but they do find that those who played a game perceived the insurance as fairer. Panda et al. (2015) explicitly compare different kinds of training. They find that a game, called treasure-pot, in which exposure to risk is simulated and players can protect against risk by pooling resources, was most successful as judged by the participants. A high proportion of participants enjoyed the activity and found it useful.

How can we explain this mixed evidence regarding the effect of financial literacy training? Since the effect varies depending on whether we deal with complex index-based insurances or with simpler health insurance schemes, training programs do seem to succeed in improving the understanding of the technicalities of insurance. None the less, three important points deserve to be made here. First, knowledge is not necessarily sufficient to translate into an effective purchase of the insurance product. For instance, a lack of trust in the insurer or cultural beliefs may interfere with the benefits of financial training. Second, the aforementioned studies measure the short run effects on demand. Yet, Platteau & Ontiveros (2013) show that understanding is also a crucial factor in contract renewal. Training in financial literacy, possibly coupled with a good follow-up, could therefore have substantial effects in the longer run. Third, a deeper understanding about the logic and the concept of insurance might be necessary, and it is possible that current training methods do not succeed in raising this kind of understanding. While it is unclear whether training in financial literacy can achieve its purpose, there is definitely scope for current training methods to focus less on the technicalities of the insurance product and more on a broader understanding of the underlying concept or mechanism. The absence of effect of training in financial literacy on the demand for simpler (health) insurance contracts might however indicate that understanding is not a main barrier to demand for those types of insurances. It is nevertheless noteworthy that Platteau & Ontiveros (2013) have studied a microinsurance program in India and, yet, they are able to conclude that lack of understanding is one of the key factors impeding insurance renewal.

3.2.3 Influence from peers

Peers do have an important influence on the decision to purchase insurance. This is suggested by Patankar (2011) and Giné et al. (2008) in whose studies insurance purchases by people close to each other are positively correlated. Similarly, Platteau & Ontiveros (2013) find that members of self-help groups are more likely to renew insurances, even after controlling for the level of information which these group members have. Such correlations should however be interpreted cautiously as it can perhaps be expected that friends and members of the same group, who can be "similar" in some unobserved ways, exhibit similar purchasing behaviour.

Three randomized control trials provide stronger evidence that, by spreading information about the insurance, peers can increase the likelihood that insurance is purchased. Karlan et al. (2014) and Cole et al. (2014) note that take-up is strongly influenced by the payout experience of peers. Similarly, Giné et al. (2014), find that the comics they distributed for financial literacy training only have a substantial effect when many in the village receive them. While this could be explained by the spreading of information among peers, it could also be that uptake decisions are spread among peers. However, belonging to a village where many discount vouchers have been distributed and take-up is thus higher, seems to have only a modest effect on individual uptake decisions. As such, the information hypothesis seems the most likely.

The results of Cai et al. (2015) corroborate this finding. They conducted an experiment in which people are assigned to an intensive or a simple training session. Moreover, this is done in two phases, a few days apart, with different individuals in each phase. They find that people who have friends participating in the intensive sessions are more likely to buy the insurance and more informed about it. Moreover, only 9% of people are aware of their friends' uptake decisions. It seems therefore that it is the increased knowledge through peers that causes increased take-up. Nonetheless, when participants in the second phase are explicitly informed about their friends' take-up decisions in the first phase, they are more likely to purchase when their friends purchased. Hence, the actual decision to purchase insurance might also influence the decision of others, although, at least in the short run, information about purchasing decisions might not spread as easily as information about the insurance itself.

Aware of the importance of this peer effect, insurance providers may also pay attention to the agents whom they put in charge of the distribution of the product. In this respect, Giné et al. (2008) used the village networks to disseminate the information about the product. They performed a more intensive marketing of the insurance product in the direction of selected village opinion leaders and asked them to help publicize the insurance product. They later came back to sell the policy. Participation is 8 percentage points higher among members of the local council (Gram Panchayat) and somewhat superior for those who are connected to other village networks. Thus, a network effect seems to be at play. The authors, however, acknowledge that, since marketing intensity is omitted in their regression estimates, the strength of the results is dampened. Networks have also been exploited by Dercon et al. (2011) who curiously find that a peer referral treatment has counterproductive effects in the sense that the tea centres in which this intervention was implemented were less likely to purchase the insurance policy. Indeed, take-up dropped from 13% to 6%, presumably because of its similarity to a pyramid scheme that broke trust.

Finally, going beyond the simple fact that peer effects may increase (or decrease) demand, Platteau & Ontiveros (2013), find that peers may substitute for a poor understanding of insurance. As we mentioned before, people in their study drop out of the insurance following a lack of payouts except if they understand the insurance. The same conditional effect is actually observed when they have a peer who renewed the insurance. In other words, both the presence of peers who renewed their insurance and a good grasp of the insurance concept have the effect of mitigating the negative impact of a lack of payouts on renewal behaviour. It is as though trust in peers acts as a substitute for a solid understanding of insurance. Even if one does not understand the insurance, having a peer that does is sufficient to make correct decisions.

3.2.4 Attitudes towards risk

Risk aversion

Under expected utility theory, risk aversion is the reason why insurance is valuable: a risk-neutral or risk-seeking individual should not purchase insurance, even when it is actuarially fair. As a consequence, demand for insurance should be higher for the more risk-averse individuals. Several studies, however, have found that risk-aversion can be negatively, and quite strongly, correlated with the demand for insurance (Cole et al., 2013; Giné et al., 2008; Giné & Yang, 2009; Dercon et al., 2011).

One explanation for this rather counter-intuitive result can be detected by looking more closely at the possibility that the insurance may fail to payout in case of a loss. This can happen if the insurer defaults (Doherty & Schlesinger, 1990), if the client distrusts the insurer to payout when required (Dercon et al., 2011), or, in fact, in about any insurance scheme to the extent that contracts almost never perfectly cover the client for every potential loss. The latter problem is especially important for index-based insurances (Clarke, 2011a). Indeed, when insurance payouts are based on an index such as the weather for an agricultural insurance, there is always the risk, called basis risk, that a farmer suffers a loss even though the weather was good, so that no payout is given.

In a situation where an insurance might fail to payout, the insurance becomes risky in itself. As a matter of fact, someone who purchased an insurance can end up in a situation in which he paid the insurance premium, suffers a loss and does not receive a payout. This scenario is worse than any attainable situation had he not purchased insurance and is therefore particularly unattractive for very risk-averse people. For this reason, when risk aversion increases, demand for insurance might first increase but it will eventually decrease (Clarke, 2011a; Dercon et al., 2011; Doherty & Schlesinger, 1990). A lack of trust in the insurance, in which case the insured also believes that he might not get a payout when needed, is thus one way in which insurance can become unattractive for the most risk averse individuals. Dercon et al. (2011) look in detail at the interaction between trust and risk aversion. They find that, controlling for trust, slightly increasing risk aversion for risk-loving individuals seems to have a positive effect on demand, but this effect becomes negative as agents become more and more risk averse, that is, extreme risk-aversion seems to decrease the likelihood to purchase the insurance. In line with these findings, Giné et al. (2008) find that the negative effect of risk-aversion on demand is mostly concentrated among people who do not know the micro-finance institution selling the insurance, and these are exactly the individuals who are expected to trust the insurer the least. Ambiguity aversion provides another explanation as to why risk aversion can decrease demand for insurance. Ambiguity about the insurance, and aversion to this ambiguity, can make the most risk averse individuals dislike the insurance. Bryan (2013), revisiting the data of Giné & Yang (2009), finds that the negative effect of risk aversion on demand is driven by ambiguity-averse individuals; demand from non-ambiguity averse individuals is increasing with risk aversion, as standard theory would suggest. In conclusion, there is evidence that risk aversion can limit the demand for insurance, which is a source of concern. When insurance is considered risky in itself, its usefulness as a risk-coping instrument is limited, and the most risk averse individuals who should benefit most from insurance do not even purchase it. Nonetheless, by reducing ambiguity about the product and increasing trust in the insurer, it seems possible to retrieve the basic result that insurance is most valued, and most often purchased, by individuals with relatively strong risk aversion.

The effect of past shocks

An individual's evaluation of an insurance contract does not only depend on his level of risk-aversion, but also on his perception of the risks he faces. In this sense, the frequency and intensity of past shocks may have a strong influence on the perceptions of risks, and hence, affect the demand for insurance products. Nonetheless, Cole et al. (2014) and Stein (2014) do not find any clear evidence that having experienced a weather shock increases the uptake of weather-based insurance.

In many instances, it is actually difficult to disentangle the exact reasons why the experience of a shock might change demand for insurance. Both a change in the perception of risk as well as the consequences of the shock could have an influence on demand. Cai & Song (2013) provide a perhaps surprising result: the experience of hypothetical shocks in a repeated insurance game has a strong positive effect on the demand for real insurance. In fact, this effect is even stronger than the effect of actually experiencing adverse events. On the other hand, Galarza & Carter (2011) suspect a judgement bias leading to the opposite effect in the project choices made by farmers in their Indian sample. When one of them suffers a loss several times in a row, he seems to be tempted to believe that chance will turn and that bad luck will not happen again in the next cropping season. He would thus underestimate the autocorrelation in the series of bad covariate shocks. A so-called "hot-hand effect"[4] might thus be at play in their sample, which drives farmers who experienced many shocks to opt for riskier projects.

In short, the experience of past hazards does not always have a clear impact on people's perception of risk, which remains a quite subjective matter.

[4]This "hot-hand effect" was first identified by Gilovich et al. (1985) who observed that basketball fans judged that a player's chances of hitting a shot was greater following a successful shot than a miss.

3.3 Low demand arising from supply deficiencies

When explaining low demand for insurance, the focus is usually on the characteristics of people that we discussed in the previous section: lack of education and understanding, particular behavioural patterns and other cognitive problems – hence the focus on trainings in financial literacy to improve demand. However, providing high quality insurance in developing countries, where transaction costs are high and good data is sparse, is difficult. In this section, we discuss the different supply deficiencies that might reduce demand.

3.3.1 Price

A first, commonly heard, explanation of low demand is that insurance is simply too expensive for the poor. To recover transaction costs on these small insurances, the commercial price of a micro-insurance is indeed substantially above the actuarially fair price (a mark-up of 50% is no exception). In fact, Clarke (2011a) argues that the price of many unsubsidized index-based insurances is so high that many expected utility maximizers are better off not purchasing insurance (this is also because of basis risk, see Section 3.3). Moreover, several studies which randomized the price of the insurance find a reasonably big price elasticity of demand, ranging from 0.44 to 1.1 (Cole et al., 2013; Dercon et al., 2011; Karlan et al., 2014). That prices are high is not the only explanation for low demand. Even when prices are significantly below actuarially fair prices, and any risk-averse expected utility maximizer should in principle buy it, bothCole et al. (2013) and Karlan et al. (2014) still observe less than 50% take-up rates. Likewise, although Bonan et al. (2012) and Thornton et al. (2010) offer health insurance for free during an initial period, only around 30% take-up was achieved in their experiments.

We may therefore conclude that, while the price seems to have a substantial impact on the willingness to buy insurance, in itself, a low price is not enough to obtain a high demand.

3.3.2 Transaction costs for the client

In addition to transaction costs borne by the insurer, there are also transaction costs on the side of the insured which often implicitly increase the price of the insurance substantially. Examples are the difficulty of purchasing or renewing the insurance, the complexity of filing a claim, and the difficulty with which premiums can be paid and payouts received. The evidence suggests that these seemingly small costs can have important effects on demand. In Thornton et al. (2010), the enrolment procedure for the health insurance they offer in Nicaragua normally requires about a day of work to complete. When, instead, they allow market vendors to sign-up directly at their market stall, uptake is about 30 percentage points higher. Similarly, Capuno et al. (2014) show that a comprehensive package of measures (including a 50% discount, SMS reminders to enrol and an information package) to boost health insurance demand in the Philippines, had only a small effect (5 percentage points) on demand. When, instead, they reduced transaction costs by offering help to fill in the forms and bring them to the agency's office, demand increased by 36 percentage points. These are very significant increases.

3.3.3 Basis risk

As mentioned before, many insurance contracts, and in particular index-insurance, come with basis-risk. For instance, a weather-index insurance may fail to indemnify a farmer's drought because the weather station, which is not located on the farmer's field, did not detect the drought.

We mentioned above Clarke's point that many unsubsidized weather-index insurances should not be purchased by any expected utility maximizer, and the reason for this is the high level of basis risk in these products. In another setting, Jensen et al. (2014a) investigate the level of basis risk of a well-designed and well-functioning index-insurance to cover lifestock mortality in Kenya, in a setting were most risk is perceived to be covariate and in principle insurable by an index insurance. The insurance reduces exposure to covariate risk by 62 percent, but to overall risk only by about 30 percent because idiosyncratic risk was unexpectedly high. Nonetheless, they find that a majority of pastoralists, but by no means all of them, would be better off buying the insurance at commercial rates. Whether the level of basis risk is low enough to justify selling the insurance thus depends on the insurance. What is clear, however, is that basis risk, by reducing the value of the insurance, can also reduce demand for it.

Several studies provide evidence about the importance of basis risk. Giné et al. (2008), studying demand for an index-insurance, argue that those farmers who produce the crops for which the policy is designed suffer from less basis risk. As these farmers are more likely to adopt the insurance, they interpret this as a negative effect of basis risk on demand. An arguably cleaner test is done by using the distance to a weather station as a measure of basis risk in weather-index insurances. Mobarak & Rosenzweig (2013), who randomly locate such weather stations, find a negative effect of basis risk on demand. Finally, instead of such proxies, Jensen et al. (2014b) use the actual level of basis risk faced by purchasers of the aforementioned livestock insurance. They find that areas where there is more insurable covariate risk (and thus less basis risk) are about 30 percent more likely to purchase the insurance.

There are different explanations for the reduced demand when basis risk is high. Clearly, basis risk reduces the value of the insurance and should thus negatively affect demand. However, as discussed before, different behavioural models that seem to explain insurance demand - including ambiguity aversion, prospect theory and uncertainty aversion – also point to a negative effect of basis risk on demand. For this reason, Carter et al. (2015) argue that reducing basis risk should be a key concern when developing insurance: It increases its value while improving the chances that it is actually bought.

In order to minimize basis risk, it is important to ensure a high correlation between the index and the losses suffered (Molini et al., 2008). In this respect, Leblois & Quirion (2013) show that different indices have widely different performances. For instance, Elabed et al. (2013) and Carter et al. (2007) argue that, for a given area, the use of area yields rather than weather data for agricultural index-insurances can significantly reduce basis risk. While an area yield insurance is very effective in reducing basis risk, it is not without problems. Ideally, payouts should be based on average yields of a small area, such as the village of the farmer. This might however introduce moral hazard: what if farmers in the village collude to reduce yields to obtain payouts? For this reason, Elabed et al. (2013) propose a multi-scale insurance contract where payouts are based on average yields at multiple levels, for instance at the village and the regional levels. They show that such a contract can address concerns about moral hazard, while still substantially reducing basis risk.

3.3.4 Contract design

Even if insuring a particular risk (health problems, failed harvest) is very valuable to people, the actual value of the insurance will still depend, to a large extent, on the contract design. When an agricultural insurance comes with a lot of basis risk or a health insurance only provides access to bad health services, the insurance will not be valuable. Similarly, many specifics of the contract, such as its simplicity or the frequency of payouts, also matter. We now discuss these different aspects of contract design.

A first aspect of the contract design is the frequency with which the insurance pays out. For instance, a health insurance can cover many or only few health problems and an agricultural insurance can cover small or only catastrophic losses. If frequency of payments is too low, a lack of payouts may jeopardize peoples' trust in the new insurance product. Yet, more frequent payouts imply a higher cost of the premiums which, in turn, may reduce demand for the insurance product. Nonetheless, Norton et al. (2014) find strong evidence that people prefer a contract with more frequent payouts. Another possibility is to incorporate different levels of payout Elabed et al. (2013). One would then obtain a partial compensation quite frequently whereas full compensation would only be triggered when losses are very big. Moreover, individuals receiving a payout are also universally more likely to renew their contract (Cole et al., 2014; Fitzpatrick et al., 2011; Dong et al., 2009; Platteau & Ontiveros, 2013; Stein, 2014), which gives an additional reason for preferring contracts with frequent payouts.

Another important challenge of the contract design is to find a subtle mix between simplicity and flexibility. On the one hand, a complex contract, even when perfectly tailored to the clients' needs, may be badly understood and therefore reduce the take-up or renewal rate. On the other hand, a simple but rigid contract may fail to meet the needs of the subscribers. Hill et al. (2011) offer an interesting alternative by tailoring the insurance to the client's specific needs through a combination of several simple contracts. They offer different weather securities that pay out a fixed amount if a specified event comes true. More specifically, the events may be monthly rainfall deficits with the farmers choosing for which months they want to insure against deficit rainfall. These "mini-contracts" offer them the ability to choose the type and number of securities, depending on their crop portfolio and production practices in a given year. Each contract is thus personalized. At the same time, the complexity of the product is increased. Nonetheless, as take-up in their study was quite reasonable, the complexity did not seem to be a substantial barrier to adoption.

Finally, the modalities of premium payments can be important. This is an issue for health insurances especially since, in an effort to limit adverse selection, the entire family is often required to enrol at once. In such cases, the total premium can be high. People therefore seem to prefer paying the premium in different instalments and paying bigger parts of the premium when more money is available, for instance just after the harvest (De Allegri et al., 2006).

3.3.5 Quality of services

When the insurance provides services rather than money, the quality of these services obviously affects the value of the insurance. For health insurance, the package mostly covers services in a designated health centre, and the (perceived) lack of quality of this centre is often identified as one of the most important impediments to the take-up of health insurance (Criel & Waelkens, 2003; Basaza et al., 2008; De Allegri et al., 2006). Additionally, the distance to the health facility can also matter (Schneider & Diop, 2004). De Allegri et al. (2006) warn, however, that people have heterogeneous preferences over health centres and do not necessarily prefer to be assigned to the closest one.

The quality of the services is a factor which people can better evaluate after having experienced the insurance and should thus affect renewal more than initial demand. Dong et al. (2009) indeed find that perception about the quality of the health centre is an important factor underlying the decision to renew. The authors report that disliking the behaviour of the medical staff is the second most cited reason for not renewing the insurance, only preceded by a concern about affordability. Yet, as with all factors affecting demand, also using high quality health centres is not, in itself, enough to ensure a high demand and renewal: Platteau & Ontiveros (2013) argue that in the dysfunctional insurance they study the health centres were actually carefully selected by the NGO which organized the micro-insurance scheme.

3.3.6 Information

A more basic lack of information about the necessary procedures and the administrative burden is another reason for clients not to buy or renew their insurance. Both Fitzpatrick et al. (2011) and Platteau & Ontiveros (2013) observe that many individuals who dropped out of a health insurance scheme reported not knowing where to make payments. While these examples may only point at ill-managed schemes, they do indicate that the effective transmission of information should not be overlooked in implementing an insurance scheme. More generally, Platteau & Ontiveros (2013) also find that a lack of information has an important negative effect on renewal. Making people well-informed is therefore crucial if one wishes to build a sustainable insurance scheme.

However, being informed about the insurance is, in itself, not sufficient for a sustained demand. An approach that is sometimes advocated is to subsidize insurance for an initial period and hope for a sustained demand after the removal of the subsidy. The idea is that the insured will then have acquired experience and knowledge about the insurance. Fitzpatrick et al. (2011) test this intuition with a health insurance product in Nicaragua. They find that, while a strong subsidy significantly increases initial take-up, many of the clients drop out after its expiration.

This confirms the intuition that those who purchase the insurance only because it is subsidized are least enthusiastic about the product. Even though overall take-up once all subsidies are ended is somewhat higher among those who initially received the subsidies than among those who did not, the results do not support the idea of using initial subsidies as a cost-effective way of increasing coverage. This finding is consistent with the work of Thornton et al. (2010), also conducted in Nicaragua, where less than 10% of the enrolled clients decided to renew their subscription after one year and after the expiration of the subsidies. Their results also indicate that the clients who received the highest subsidies were least likely to renew their contract.

3.4 Trust

By purchasing an insurance, an individual accepts to disburse a regular premium in return for an uncertain future payout. In contrast to micro-credit, one gives money first and receives it only later. One therefore needs to trust the insurer to actually payout when required, and the level of trust is expected to have an influence on demand. Following Patt et al. (2009), let us distinguish between three levels of trust: trust in the product itself, trust in the institution involved, and the degree of interpersonal trust among individuals.

First, trust in the product itself is closely related to the understanding of the product. In order to maintain trust in the product, a potential client must clearly see that by paying a premium, he (she) will be able to make choices free from the fear of losing his (her) investment in case of an adverse shock. Platteau & Ontiveros (2013) indeed find that insurance understanding and trust in the insurer are highly correlated and often difficult to disentangle from each other. In this respect, Patt et al. (2009) discuss the ability of experimental research to build trust in the product. They assemble considerable evidence from case studies to show that, through participatory methods, farmers are able to learn how the insurance contract works, and how to explain it accurately to others. In addition to enhancing understanding of insurance, they argue that such games are also valuable in building trust in the product. Another key property of the insurance that may affect trust is basis risk. The insurance might be considered of low quality and not trustworthy when losses are frequently incurred while no indemnity payment is received from the index insurance.

The second dimension of trust - trust in the institutions involved - is the dimension of trust most discussed in the literature. Several factors which influence trust in the institution can be identified: experience with the institution, the involvement of known and trusted individuals, trust in the management of the institution, and other external factors. The potential trust-building role of experience with the institution that delivers the insurance is highlighted by the fact that experiencing a payout seems to be an important factor in the decision to renew: receiving a payout is systematically associated with substantially higher renewal rates (Cole et al., 2014; Fitzpatrick et al., 2011; Dong et al., 2009; Platteau & Ontiveros, 2013; Stein, 2014). A difficulty in assessing whether these payouts actually cause higher renewal rates, is that they might correlate with other factors influencing renewal. The shock that triggers the payout can indeed in itself influence the decision to renew. Similarly, less healthy individuals receive more payouts from a health insurance (now and in the future) and should therefore value insurance more. However, Platteau & Ontiveros (2013) find that, even when controlling for individuals' current health status, the effective use of a health insurance in case of illness increases the likelihood of renewal. Similarly, Stein (2014) and Cole et al. (2014) provide evidence that weather shocks do not drive take-up decisions, and it is therefore the actual occurrence of payouts which increases contract renewal.

The most likely explanation of the positive effect of payouts on renewals is that they enhance trust in the insurer, but other explanations have also been proposed. Stein (2014) argues that increased renewal can be ascribed to loss-aversion: without payout, the payment of the premium is seen as a loss, while with a payout, the premium payment is seen as reducing the payout, and thus reducing a gain. As the loss is felt more strongly than the reduced gain, a payout thus makes the payment of the premium less painful and increases the probability of renewing the contract. Cole et al. (2014) as well as Karlan et al. (2014) however find that uptake also increases when one's peers receive a payout, which is not consistent with this explanation, but rather points to increased trust following the payout. Moreover, Cole et al (2013a) find direct evidence for the trust explanation: people who have received a payout report to have substantially (though not significantly) more trust in the insurer. Payouts thus seem to increase renewal by increasing trust in the insurer.

Additionally, it is important to note that a general distrust in financial institutions, or a bad experience with other institutions, can decrease trust in new insurers. Basaza et al. (2008), for instance, claim that it took two years to overcome distrust, caused by bad previous experience, thanks to positive experience with a new insurance. Finally, Patt et al. (2009) report that farmers say that they put their greatest trust in organizations that they themselves are members of and that, in general, trust increases with experience with the organization. Similarly, members of self-help groups have been found by Platteau & Ontiveros (2013) to be more likely not only to subscribe but also to renew their health micro-insurance contract.

The involvement of known and trusted individuals in the insurance is also shown to have a substantial influence on the demand for insurance. Giné et al. (2008) find that members of borewell user associations in India, who are more likely to know the insurance vendor personally, are 37 percentage points more likely to buy the insurance contract, or 7 percentage points more likely when they restrict the sample to existing customers of the MFI. Those figures are however to be considered with caution since the authors could not control for the marketing intensity of the insurance product in the direction of village opinion leaders and existing customers, which may partly account for this effect. Similarly, Cole et al. (2013) find a strong and significant effect of introducing an insurance educator into the visited households by a local trusted agent from BASIX. The interpretation that such an agent enhances trust is strengthened by the fact that, amongst households familiar with the BASIX microfinance institution there is a 10 percentage point increase when a BASIX agent does the introduction, while there is no significant effect for those who are not familiar with BASIX. Indeed, for the latter, an unknown agent would not be expected to enhance trust.

Dong et al. (2009) identify trust in the management of the community health scheme as an important factor influencing households' probability of enrolling. The involvement of known and trusted individuals in such a scheme could thus have a positive effect on demand. De Allegri et al. (2006), however, note that although people like to have representatives of the community-health scheme at the local level, they prefer the money to be managed outside of the community (such concerns can be well-founded; see Vanderwalle (2015) for an example of elite capture of local resources). When there is a lack of trust in the community management, the involvement of people representing it can thus have a negative effect on take-up.

Finally, external factors can have an influence on trust placed in the insurer. Schneider & Diop (2004) and Patankar (2011) point at the potential trust-building role of legal and institutional support. Moreover, Patankar (2011), studying an index-insurance, indicates that people would place more trust in the measurement of the index if it is certified by the government.

The last aspect of trust which can have an influence on insurance uptake is interpersonal trust. If an individual does not trust his circle of friends and neighbours, he may be, in general, less trusting in others.

Therefore, he may be suspicious when requested to take part in a formal risk pooling organization. Patt et al. (2009) argue that trust in other people responds to various social and cultural factors. This aspect of trust is often measured through variants of the trust game initiated by Berg et al. (1995). Here again, one should be aware of what is captured by the trust measure. Some games try not to confound the trust effect with altruistic motivations, by first playing a dictator game[5]. It is also possible to distinguish trust from fairness considerations. When comparing empirical studies that probe into the issue of trust, we therefore need to be precise about the content of the trust measure. The trust game is, for instance, used in the study of Dercon et al. (2011) regarding a composite health insurance product in Kenya, and the authors show that low interpersonal trust levels correlate negatively with insurance uptake.

3.5 Insurance substitutes

Micro-insurance is not the only option to mitigate and cope with risks. Other tools such as credit, precautionary savings, informal risk-sharing agreements, and self-insurance strategies also offer (partial) protection against risks. By providing a substitute for insurance, the availability of these tools can reduce the demand for micro-insurance.

The aim of the present section is therefore to investigate how these alternative strategies perform in insuring poor people against adverse shocks. In particular, it will point at the benefits and limitations of these tools as risk-coping devices. Moreover, as they can affect demand for insurance beyond providing a simple substitute, this section will also highlight the different interactions between these risk-coping tools and the demand for insurance.

3.5.1 Credit and savings

Liquidity constraints

Access to credit can affect insurance demand in several ways. First, imperfect credit markets may prevent those who face liquidity constraints from taking up insurance. Most evidence is consistent with the fact that liquidity constraints are a significant barrier to the take-up of insurance.

[5] In this game (Kahneman & Tversky, 1979), players receive a monetary endowment and are asked to send all, part or nothing of it to another player, whose identity is not revealed. The higher the amount sent, the more altruistic the agent is considered.

Wealthier individuals are typically more likely to buy insurance(Cole et al., 2013; Gaurav et al., 2011; Giné & Yang, 2009). Additionally, handing out money right before the purchase decision, which relaxes potential liquidity constraints, significantly increases take-up (Cole et al., 2013). Of course, take-up might increase simply because money is given and people reciprocate by doing the perceived right thing: buying insurance. Cole et al. (2013), however, argue that liquidity constraints matter because they observe that the big endowment has a larger effect on poorer individuals, for whom liquidity constraints are more likely to be binding.

Karlan et al. (2014), however, reach opposite conclusions. They find that giving cash grants increase insurance take-up (irrespective of its price) in a credit constrained environment. They argue that the main mechanism driving the higher take-up is the signal cash grants give to the recipients (and their peers) rather than the alleviation of a liquidity problem. Indeed, they bring evidence to discard the wealth effect and conclude that people either reciprocate the benefit of a grant by buying the insurance or that they have increased trust in the insurance when they have received a cash grant.

Remittances are another way of relaxing liquidity constraints. Yet, their impact on the decision to purchase the insurance is ambiguous. Indeed, beyond increasing income, remittances also provide self-insurance, and thereby, substitute for micro-insurance. Crayen et al. (2013) show that remittances, after controlling for income, act as a substitute with respect to formal funeral insurance in a context where the budget constraint is binding. Likewise, Giesbert et al. (2011) show that those who have received remittances in the past, and could thus potentially receive remittances in the future to cope with a shock, are less likely to take up insurance.

Credit as substitute for insurance

An additional way in which credit can affect insurance demand is through its role of consumption smoothing, thus allowing to cope partially with the consequences of an adverse shock. Nonetheless, credit is a highly imperfect insurance instrument for several reasons. First, although credit allows to spread the effects of a shock over time, a big shock will still leave one worse off than in the case the shock had not occurred. Since a risk-averse individual prefers to perfectly smooth consumption, both over time and over states of nature, this is suboptimal. Second, the likelihood to be granted a loan for insurance purposes could actually be lower than normal, precisely because it is requested when the individual is most vulnerable and thus least likely to be able to repay. Finally, the serious consequences in case of default increase the risk faced by resourceless individuals.

A more hybrid transaction, half-way between credit and insurance, has been identified by Platteau & Abraham (1987) as "quasi-credit". Incidentally, by letting repayment depend on the situation of both borrower and lender - transactions are personalized with the possibility to renegotiate reimbursement following shocks and loans are often made without interests - quasi-credit suffers less from the aforementioned limitations. It is even debatable whether it is best described as credit or as risk-sharing agreement (Platteau & Abraham, 1987; Udry, 1990).

Whether access to credit is positively or negatively correlated with the demand for insurance is unclear. Giné et al. (2008) find that those who are credit-rationed, that is, those who want a loan but are not able to get one, are significantly less likely to purchase insurance. As for Hill et al. (2013), they observe that those who deem themselves capable of gathering a certain sum of money in a week if necessary, are less likely to purchase insurance. The evidence is therefore rather inconclusive.

Savings

Savings, in cash or in the form of marketable assets, allow to cope with the consequences of risk in a way very similar to credit. As a (self-)insurance device, savings also face the limitation that shocks are spread over time rather than over states of nature. In addition, the size of the shock which savings can help to overcome is bounded by the amount of the available savings. This is a major limitation, especially when several shocks occur in a short period of time. Another drawback of precautionary savings is that it has a cost, namely the investments potentially forgone.

Interlinking of insurance, credit and savings

Besides the potential of credit and savings to substitute for insurance in coping with risks, there may also exist complementarities, for example by interlinking credit and insurance. Indeed, the possibility of default on a loan creates risk for both the bank, who might lose part of its money, and the customer, who might lose his (her) collateral. In this case, offering credit together with an insurance has several advantages (Carter et al., 2011): the bank, facing a lower default rate, can charge a lower interest rate and the customer, facing less risk and a lower interest rate, has two reasons to demand more credit. Moreover, as the credit is linked to insurance, a higher demand for credit will automatically increase insurance take-up. However, this may be dangerous if clients are not aware of what they actually buy and are, somehow, forced to take the insurance along with the loan. This approach would then go against all efforts to make the product understandable. Nonetheless, when a farmer faces a profitable but risky investment opportunity - such as the purchase of high-yielding seeds - an interlinked credit and insurance contract could be the only tool available to make the investment attractive, and would thus be especially valuable (Carter et al., 2011).

Several studies investigate the effect of jointly offering credit and insurance. Giné & Yang (2009), when offering credit to finance investment for new seeds, randomly oblige one group of customers to jointly take-up an index-based insurance (at market price) to mitigate the risk inherent in taking a credit for the adoption of a new technology. Similarly, Banerjee et al. (2014) study a situation where a micro-credit organisation obliges clients to take an obligatory health insurance when taking a loan. Although the expectation is that insurance makes the adoption of credit less risky and thus more attractive, both papers find the opposite effect: Giné & Yang (2009) find that the take-up decreases by 13 percentage points while Banerjee et al. (2014) find a decrease of 16 percentage points. These results are driven by a low demand for insurance, which was well warranted given its low quality. In the case of Giné & Yang (2009), there was limited liability for reimbursing the credit and clients were thus asked to insure against a risk for which they were already partially insured. In the case of Banerjee et al. (2014), clients did not receive information about the insurance and very few actually ended up using it. In the light of these circumstances, a low demand for insurance may not be surprising, but it does teach us another important lesson: Making insurance mandatory should only be considered when it has proven to be well-functioning and provide effective protection so that it is truly valuable.

Using a lab experiment, which allows to simulate a situation with a well-functioning and voluntary insurance, Galarza & Carter (2011) arrive at a more optimistic conclusion about the interlinking of credit and insurance. They find that 60% of people who do not take up credit in the absence of insurance, do take credit when it is secured by an insurance. This indicates that demand for insurance in such a setting can be high, and that interlinking credit and insurance can increase demand for credit and motivate the purchase of insurance. This study therefore goes against the conclusions reached by Giné & Yang (2009).

There can also be complementarities between savings and insurance. As we know, when insurance comes with much basis risk - as is the case for many index insurances - its value decreases substantially. In this situation, savings are especially useful to cope with situations in which a shock occurs but the insurance fails to payout. For this reason, savings limit the negative effects of insurance and improved access to savings could effectively increase demand for insurance with much basis risk (Clarke et al., 2012). In an experiment to assess whether linking savings and insurance could increase uptake, Stein & Tobacman (2012) find that people in fact prefer a pure insurance or pure savings product over a mixture of the two. In another experiment, Clarke et al. (2012) offer people the opportunity to allocate money to different index and health insurances. When a "group savings" scheme is added to the options available, demand for index-insurance could increase due to the complementarities between savings and insurance with basis risk. They nevertheless find that the introduction of savings does not significantly change demand for index-insurance, suggesting that these complementarities are not very important. Yet, people do indicate that savings are an important part of their risk-management strategy since they allocate a good part of their money to the group savings scheme.

3.5.2 Informal risk-sharing networks

Besides the aforementioned individual risk-coping strategies, people can also engage in informal insurance mechanisms in which they mutually provide help to each other in times of need. Such informal arrangements allow to cope with unexpected health or schooling expenses, necessary disbursements for funerals or other important ceremonies, among other things. They can take the form of actual risk-sharing groups, such as funeral societies, or of flexible transfers between family and friends (Fafchamps & Lund, 2003). When efficient, such informal risk-pooling agreements can crowd-out insurance (Arnott & Stiglitz, 1991). Compared to micro-insurance they present informational advantages since members can monitor each other more efficiently, but they have the drawback to represent only a limited pool of risks. An “insurance dilemma” thus arises (Platteau, 1991). The limited pool of risk not only weakens their ability to deal with large covariate shocks, but also threatens their stability if multiple idiosyncratic shocks occur in a short period of time.

Informal risk-sharing agreements can therefore influence demand for insurance in different ways. On the one hand, insurance and informal risk-sharing can be considered as substitutes. Yet, if people start leaving informal groups because of the offer of micro-insurance, the risk pool of the risk-sharing group is reduced, which makes them less effective in coping with risk (Fafchamps, 1992). For this reason, participation in formal micro-insurance schemes might be discouraged by group members. Moreover, membership of a risk-sharing group could delay the adoption of insurance when the group cannot be left immediately. When, for instance, one has received or be given more than the others in a funeral society, it might not be permitted or desirable to leave right away.

On the other hand, one can also consider the take-up of insurance as the adoption of an innovation. Here, it is typically the more educated, wealthier, and less risk-averse individuals who are early adopters (Rogers, 1995). However, those who are more accustomed to similar technologies are more likely to adopt the innovation. In this case, informal risk-sharing agreements might be such a similar technology and membership might thus encourage adoption.

Complementarities may exist between informal risk-sharing and insurance in much the same way as they exist between savings and insurance. First, risk-sharing groups can be an especially efficient way of spreading information about insurance, thereby promoting its uptake. In addition, because an insured individual is better protected against shocks and is more able to help other members of the group, he is of more value to them. On this count, insurance participation should be encouraged by group members. Conversely, when an insurance comes with much basis risk and a risk-sharing agreement succeeds in providing coverage when the insurance fails to pay out, improved risk-sharing attenuates the negative effect of the basis risk and can thereby increase demand for insurance. Mobarak & Rosenzweig (2013) find that when they cover similar types of risks, informal risk-sharing is a substitute for micro-insurance. Such substitution effects should not be at play when the informal insurance covers other risks, and indeed they do not find an effect of the extent of informal coverage of idiosyncratic risks on the demand for insurance when there is little basis risk. However, when basis risk is high, more informal coverage of idiosyncratic risks has a positive effect on demand, confirming the idea that informal coverage can attenuate the negative effects of basis risk present in micro-insurance contracts. These results thus confirm that, by exploiting complementarities with existing risk-coping strategies in the design of insurance, both the value of, and demand for, micro-insurance can increase (Clarke, 2011b).

One way to exploit such complementarities is proposed by Clarke & Dercon (2009) and (Clarke, 2011b). Instead of offering insurance to individuals, they propose to offer insurance to groups to pre-existing risk-sharing groups. These groups have many advantages over micro-insurance, such as lower transaction costs and less asymmetric information. However, their stability is also continuously threatened by a big covariate shock, or multiple idiosyncratic ones. By, for instance, providing an index-based insurance to a pre-existing risk-pooling group or by re-insuring those informal groups when multiple adverse shocks occur over a short period of time, an insurance can significantly strengthen such groups. Conversely, the group has the necessary information to spread the benefits of the insurance payouts to the members who are most in need, and it can thereby increase the value of insurance. In this way, informal risk-sharing groups and micro-insurance, instead of competing, could strengthen each other.

In a study by Dercon et al. (2014), a first attempt is made at understanding whether such complementarities can be exploited by offering insurance to pre-existing risk-sharing groups. In particular, the product sold is still an individual index-based insurance, but a group-focused training is provided to some group leaders which emphasizes the benefits of the insurance to the group and explains how basis risk can be attenuated through side-payments inside the group. As the group-focused training of the leaders leads to a 15 percentage points higher take-up than the ordinary training, people seem to be at least receptive to the idea of combining informal and formal risk-sharing, which is promising with respect to its potential to increase demand.

3.5.3 Other substitutes for insurance

In order to reduce the risk they face, and, for instance, protect themselves against a harvest failure, individuals may prefer to opt for a low return, low risk production strategy instead of a high return, high risk one. When comparing such self-insurance strategies to micro-insurance, it is important to realize that, as argued by Carter et al. (2011), these strategies are "neither actuarially fair, nor free of basis risk". Indeed, the reduction in average productivity makes self-insurance strategies costly while their failure to remove all risks amounts to the presence of basis risk. Thus, when micro-insurance allows to relax certain self-insurance strategies, the question is not whether it is too costly or carries too much basis risk, but whether it is more efficient in dealing with risk than the self-insurance strategies it replaces (Carter et al., 2011).

Finally, depending on the context, several other risk-coping instruments are available. As pointed out above, family members informally share risk through flexible transfers. Jowett (2003) thus finds that where private transfers among people are important, people are less likely to purchase health insurance.

To sum up, micro-insurance is developed in an environment where people have access to a variety of tools to cope with risk. When developing an insurance, it is thus critical to understand which risks are least covered by existing arrangements since micro-insurance is likely to be most valuable, and in highest demand, when it complements rather than substitutes for existing arrangements. As we have seen, the different existing risk-coping strategies are generally least effective for large shocks, and even more so when covariant risks are at stake. These risks are therefore the main source of the comparative advantage of micro-insurance (Clarke & Dercon, 2009). For smaller, idiosyncratic risks, micro-insurances schemes are not necessarily the best instrument to the extent that their effectiveness may be hampered by informational problems and high transaction costs. Nonetheless, since even those specific risks are only incompletely covered by existing methods, even here micro-insurance could potentially offer some additional protection.

3.6 Conclusion

Perhaps the most puzzling question regarding micro-insurance is the following: why are take-up and renewal rates so low while micro-insurance may significantly increase the protection of the poor against adverse shocks?

Paradoxically, demand seems to be negatively correlated with risk aversion whereas it should be valued as a risk-coping instrument. This observation emphasizes the current limitations of micro-insurance. Various theories offer possible explanations for the paradox. Yet, evidence is far from decisive. In particular, lack of knowledge about the nature and technical characteristics of micro-insurance products is not sufficient to account for low demand. Indeed, although literacy training appears to significantly enhance knowledge by shedding light on the specificities of complex insurance products, it is less efficient when it comes to enhancing demand. An important reason may be that people need to have a good grasp of the very notion of insurance, and not only of technical characteristics, to be able to correctly perceive its benefits. In this respect, it is revealing that many demand-related problems discussed in this paper have also been observed in the context of developed economies, the United States in particular. Kunreuther et al. (2013) have labelled these problems the "underpurchase demand-side anomaly".

Moreover, several other factors also influence the purchase decisions of individuals. As expected, demand responds to changes in the price of the product, with a higher take-up when the premium becomes more affordable. However, a low price is not enough to obtain a high demand. Thus, a wide range of evidence illustrates how a lack of trust, which may itself be caused by a lack of understanding of the insurance concept, may also constrain demand. Poor quality of the product is unanimously regarded in the empirical literature as another factor that contributes to lower demand. The evidence is, however, much more divided when it comes to the availability of credit. It is not clear whether individuals who have easy access to loans are more likely to subscribe, or not.

In the light of these considerations, what can be done to enhance demand? In which field is further research needed? In the current context, two options are conceivable. The first is to rethink the insurance contracts in order to mitigate the various problems highlighted above. One could, for instance, imagine a double-trigger contract for index insurance, which would offer frequent payouts, and, at the same time, cover big shocks. Further research in this direction is desirable. Adapting the modalities by spreading the premium payments over several months, in order to relax liquidity constraints, is another possibility. We also discussed the potential of interlinking insurance with credit or savings. However, this may also imply negative consequences if, for instance, people are not aware or do not understand the product they buy. Moreover, we need to think about reliable ways to reduce the basis risk inherent to an insurance contract. To overcome the many other factors limiting demand, innovations in micro-insurance are necessary and should be imagined. Many of the conceivable ideas, however, give rise to trade-offs between the necessary simplicity of the contract and the flexibility it offers to the client. For this reason, they are unlikely to dramatically increase demand.

Therefore, we explored the potential of informal risk-sharing arrangements as an alternative way of increasing the attractiveness of micro-insurance. We believe that more research could help better understand the possible complementarities between formal and informal practices, for instance by offering insurance via the vehicle of a pre-existing, informal group. In addition, trust in the insurance can be built in several ways: involving trusted organizations or individuals; ensuring sufficient payouts to create a positive experience; or adding government certification and regulation to avoid that defaulting insurers break trust in others. These are some promising ideas but more work is required to identify the most effective ways to enhance trust in the product itself and in the institution delivering the insurance. Finally, more attention ought to be given to the problem of low renewal rates, which has been largely disregarded so far. One useful direction for research would be to look at the long term impact of literacy training on renewal rates. Are those who attended a training session less prone to be disappointed in the absence of payout and, therefore, more likely to renew? What is the scope and is there value added in a potential follow-up of the clients?

Finally, mandatory state-directed insurance may appear as the ultimate solution to the failures of micro-insurance. But there are two major problems with state insurance. First, it may not reach people in the informal sector because the payment of the premium is not enforceable. Second, it may give rise to severe management or governance problems akin to state failure. In the end, comparing the effectiveness of mandatory state-directed insurance with that of micro-insurance comes down to assessing state versus market failures. Interest in micro-insurance was actually caused by state failure as attested by the disappointing performances and insufficient outreach of state insurance schemes. The fact that micro-insurance has not proven as effective as initially hoped for shows that the question of the respective roles of state and private insurance should be squarely put back on the research agenda.

Part II

Applied Econometrics

Chapter 4

Anchoring when measuring expectations: A methodological experiment in Burkina Faso

Abstract

We conduct an experiment on the methodology of measuring expectations. We elicit income expectations of prospective migrants in Burkina Faso planning to go and work on a gold mine. To do so, we elicit probabilities to earn different amounts of money, but we randomize these amounts: Some respondents are asked about bigger amounts than others. This seemingly irrelevant methodological change has an important effect on the responses. Median expected incomes are up to four times higher when expectations are elicited over big amounts instead of small amounts. We argue that people anchor to the proposed amounts and adjust their responses accordingly. Moreover, this undesirable effect is bigger for people with no previous experience in mining, who likely are less knowledgeable about the income distribution. These results suggest caution is needed when eliciting and using expectations data, in particular when respondents do not know the distribution well.

4.1 Introduction

Expectations are a key ingredient in many economic decisions. Indeed, many decisions involve uncertainty, and expectations about this uncertainty shape the decisions people make. As a consequence, much theoretical work in economics incorporate expectations.

Empirical work, however, has not followed suit and only very rarely addresses expectations directly. For instance, while rational expectations are an important assumption in many models, there is very little empirical work that sheds light on the validity of this assumption. Undoubtedly, the main reason for this lacuna is that expectations are only rarely measured in surveys. This, in turn, is probably a consequence of a general skepticism among economists about the ability to accurately measure people's expectations.

Recent work suggests however that this skepticism is not necessarily warranted. In a review paper on measuring expectations, Manski (2004) argues that measured expectations are generally coherent, accurate and meaningful and recommends that more surveys incorporate the measurement of expectations. Delavande et al. (2011b) argue that the same is true even in developing countries where low levels of education could complicate measuring expectations. Moreover, when comparing different elicitation methodologies, Delavande et al. (2011a) find remarkably similar results, increasing further still our confidence in these measurements.

Yet, while measured expectations usually accurately predict actual outcomes, this is not always the case when respondents are badly informed (Jensen 2010; Delavande & Kohler 2009; McKenzie et al. 2013). These badly informed people could, of course, have wrong expectations. But these result could also suggest that measured expectations are not reliable when respondents are uniformed. For this reason, Delavande et al. (2011a) call for more methodological experiments involving less well informed respondents.

In this paper, we do exactly that. We elicit income expectations of prospective migrants in Burkina Faso planning to go and work on a gold mine and randomly vary the way these expectations are elicited. Concretely, we elicit probabilities to earn different amounts of money, but we randomize these amounts: Some respondents are asked about bigger amounts than others. Testing whether responses change following this seemingly irrelevant methodological change allows to test how reliable answers are. Moreover, since some respondents have experience in mining while others do not, we can also compare the reliability of answers based on how well informed people are.

We find that answers depend strongly on the way the question is asked. Median expected incomes are up to four times higher when expectations are elicited over big amounts instead of small amounts. Moreover, this undesirable effect is bigger for people with no previous experience in mining, suggesting that less informed individuals indeed give less reliable answers.

These results thus suggest caution is needed when using these data. Clearly, we can not infer whether expectations are, for instance, too optimistic. Moreover, the biases we find are big and heterogeneous (experienced people are less affected). As such, the data measure as much experience in mining as they do expectations, and even using the data as a simple proxy for expectations can be problematic.

We argue that these results are driven by “anchoring", a psychological phenomenon by which people’s answers are affected by a seemingly irrelevant number that is presented. For instance, people’s answers to a factual questions are affected by the outcome of spinning a wheel (Tversky & Kahneman 1974). This phenomenon has been studied extensively in psychology (see Furnham & Boo (2011) for a review). Likewise, the inconsistency of answers in contingent valuation tasks (Diamond & Hausman 1994) has been repeatedly linked to anchoring effects (Green et al. 1998; O’Conor et al. 1999; Van Exel et al. 2006; Luchini & Watson 2013). Our results are perfectly in line with these literatures. As usual, we find that anchoring effects are big and that they are most pronounced on people with less knowledge about the subject. Our main contribution with respect to these literatures is that we confirm these results in a setting where multiple anchors are presented instead of just one, as is typically done in contingent valuation and psychological studies on anchoring.

The remainder of the paper is structured as follows. In the next section we describe the experiment and show the main results. In Section 4.4 we argue that the most likely explanation of these results are anchoring effects and in Section 4.5 we address the problems that such anchoring effects create. To understand when anchoring effects might arise, and what can be done to avoid them, Section 4.6 discusses the factors affecting anchoring effects. The last section concludes.

4.2 Experiment

4.2.1 Setting

The experiment was done with a sample of 112 coton farmers in South-West Burkina Faso. The experiment was part of a larger-scale project studying insurance for cotton farmers.

While the main occupation of the respondents is farming, expectations were elicited on the returns to artisinal gold mining. Following big increases in gold prices over the last decade, artisanal gold mining is a widespread activity in most of Burkina Faso. As opposed to industrial gold mining, artisanal gold mining is mainly done manually on sites (ranging from a few workers to several thousand ones) all across Burkina Faso. Throughout the study area there are active gold mines close to the villages (at a distance of a few up to 50 km) of the respondents. Work on the gold mines is mostly done during the dry, lean season when there is little agricultural work.

The sample consists of farmers who are household head and who in the month following the harvest season (the moment of the interview) were considering to go and work on a gold mine. Out of 1014 farmers interviewed, 112 were considering to go and work on a gold mine[1]. We elicited expectations about the income they were expecting to earn on the gold mine.

Income from gold mining is highly uncertain. Once arrived on a mining site, people group in small groups of about 6 and work in one pit. Gold found in the pit is shared among the group members and is converted to cash on the site itself. Most sites are illegal and all the income is kept by the team members[2]. As can be expected, returns from mining is very uncertain. Figure 4.1 shows the historical distribution of earnings for 1 month of work on the mines. The distribution is heavily skewed to the right: Few people earn big amounts, while many earn little (and about 25% earn nothing at all).

The respondents are all cotton farmers who are also prospective miners. Table 4.1 shows some of their characteristics. About 62 % had experience in mining at the moment of the interview, the others were expecting to leave for the first time. The ones with experience typically have a recent experience: 84 % of them has worked on a mine during the last year. The respondents are all household heads and are on average 38 years old. The general level of education is low: 80 % has no education, and the average number of years of education is 4.3 for the ones who did get education. 32 % can read and write in at least one language.

[1] About 28 % of these farmers actually ended up leaving for the mines. For comparison, about 5% of the other respondents ended up going to the mines.

[2] In a small number of legal sites, a "mining company" is present. However, the only role of the mining company is to guard the site, prevent conflicts and collect a small part of the miner's revenues in return. Miners are never in wage labour and their income is a direct function of the gold they find.

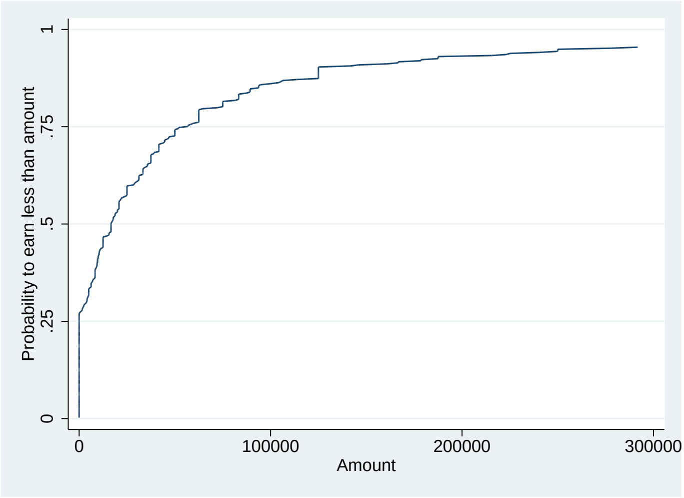


Figure 4.1: Historical distribution of earnings on a gold mine for a month of work. Total yearly earnings of miners (who worked at most 2 months on the mines in that year) are transformed in monthly earnings. We use data from the three years preceding the moment of the interview.

	N	Mean	SD	Median	Min	Max
Age	112	38.60	10.46	37.00	19.00	82.00
Has experience mining (=1 if yes)	112	0.62	0.49	1.00	0.00	1.00
Has any education (=1 if yes)	112	0.20	0.40	0.00	0.00	1.00
Nr. years education (if has education)	22	4.36	2.11	4.50	1.00	9.00
Can read and write (=1 if yes)	112	0.32	0.47	0.00	0.00	1.00

Table 4.1: Descriptive statistics for the 112 cotton farmers participating in the experiment.

4.2.2 Description of the experiment

Eliciting expectations from farmers with a low level of education is a challenge.[3] The concept of probability, and the hypothetical question it implies, are particularly difficult. As advised by Delavande et al. (2011b), we use graphical aids, which are known to work well. More concretely, we use 20 stones which must be distributed over different outcomes. Here, every stone reflects a probability mass of 5% and so the stones allow to respond to probabilistic questions by rounding to the nearest 5%[4].

Concretely, we pose the following question to elicit preferences:

> I would like you to imagine that there are 20 gold miners like you who will leave for the mining site this dry season to go and work in the pits. These stones represent the miners. [Show 20 stones] Every stone reprents one miner. I imagine that some of them will be lucky and others not. I would like to ask you what income they will have (the income does not include gifts from others). Can you distribute the stones to indicate how many of them will have a revenue of 0, 400, 1000, 4000 and 20000 (or more) after one month of work?[5]

Following this question, the enumerator would spend time with the respondent to make sure that he understood, including verifying that the answer given corresponds to his expectations (e.g., if somebody puts all stones on one amount, asking whether everybody will earn that amount and nobody the other amounts).

[3]A simple way to elicit expectations is simply to ask “how much do you expect to earn?" instead of eliciting the full income distribution as we will do. However, it is unclear what the simple expectations question actually measures: the mean, median, mode or a combination of these? Additionally, Delavande et al. (2011b) show that the answers to this question have less predictive power than the ones obtained by eliciting the entire income distribution. For these reasons, we try to elicit respondent's full income distributions.

[4]See Delavande et al. (2011a) for experiments and a discussion on the optimal number of stones to use. We followed their advice of using 20 stones.

[5]This question allows respondents to report the distribution of incomes they expect to earn. However, respondents could be very uncertain about the exact form of this income distribution. In this paper, we do not elicit such uncertainty. Indeed, this is very challenging since it combines two levels of uncertainty: the distribution itself and the uncertainty about this distribution. It could for instance be done by asking for the minimum and maximum probability to earn a certain amount. Nonetheless, the anchoring effects we will study are likely caused by uncertainty about the distribution and eliciting this uncertainty in a simple way could give useful insights into anchoring problems (see Luchini & Watson (2013) for an example in the context of contingent valuation).

Category	Amounts					N
Very low amounts	0	400	1000	4000	20000 (or more)	31
Low amounts	0	1000	2500	10000	50000 (or more)	24
High amounts	0	2000	5000	20000	100000 (or more)	33
Very high amounts	0	10000	25000	100000	500000 (or more)	24

Table 4.2: Amounts used to elicit expectations and the number of respondents for which each set of amounts was used.

To elicit the individual's expectations about his own returns, we talk about *20 gold miners like you who will leave this dry season.* Here, the key is that they are 20 miners *like you* and should thus have similar earnings. The alternative to ask about (hypothetical) realizations over the last 20 year would not have been appropriate: Gold prices as well as profitability of sites fluctuate too much over time for the past to be an accurate prediction of the future.

The experiment we conduct is to randomly vary the amounts presented in the question to elicit expectations. The amounts shown in the question are the set of "very low amounts", but we also use different sets of amounts (presented in Table 4.2). The amounts scale up linearly, with the low amounts being approximately double the very low ones, the high double the low ones and the very high amounts about 5 times the high ones. With respect to the "true" (average) distribution, the low and high amounts allow to accurately describe expectations over most of the distribution, while the very low (resp. very high) amounts allow for most precision in the left (resp. right) half of the distribution. Randomization happened at the individual level and the randomization is well-balanced.[6]

4.3 Results

The goal of the empirical analysis is to study the framing effect: When expectations are elicited over higher amounts, do people report higher expectations? Clearly, when the question is asked differently, people should give different answers. However, their answers should always reflect the same underlying distribution of their expectations. To show that there is a framing effect, we need to show that the underlying distribution that

[6]We tested equality of averages of respondents presented higher (high and very high) amounts against the ones presented lower (low and very low) amounts. We tested variables related to education, wealth and agriculture (productivity and area cultivated) and there are no significant differences.

is reported changes when the expectation question changes.

However, before addressing this question, we first need to clarify how the expectation question has been interpreted by respondents (Section 4.3.1). Based on this, we will use two methods to assess whether there are framing effects: through a simple comparison of respondents' answers (Section 4.3.2) and by fitting and comparing income distributions reported by respondents (Section 4.3.3).

4.3.1 Interpretation of the expectation question

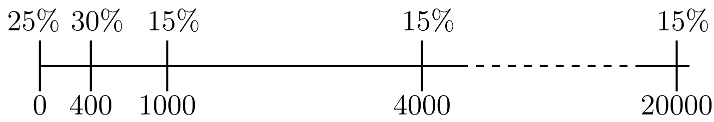
Before analyzing the data, we need to clarify how respondents might have interpreted the expectation question. We ask for the probability to earn, for example, 0, 400, 1000, 4000 or 20000 (or more). Clearly, these are not the only amounts the respondent can earn. He could equally well earn 150. However, by forcing the respondent to distribute the 20 stones (100% probability mass) over the five outcomes, we force him to consider only those outcomes. He thus needs to report probabilities to earn "about 0", "about 400", and so on.

However, there is some room for interpretation when judging what "about 0" and "about 400" exactly mean. This is illustrated in Figure 4.2. Figure 4.2(a) shows the answers of a respondent: He gives a probability of 25% to earn about 0 and a probability of 30% to earn about 400.

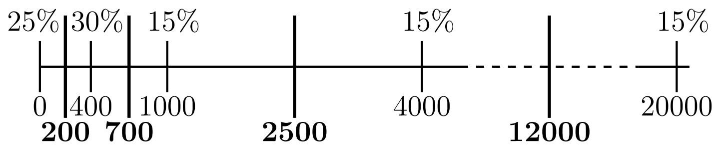
Figures 4.2 (b) and (c) show two possible interpretations of the question. In Figure (b) he assigns the probability to earn a certain amount to the proposed amount that is closest to it. For instance, the *cutoff point* between 0 and 400 is at 200. This implies that 150 is considered "about 0" and not "about 400", while earning 250 would be considered "about 400". Adopting this interpretation, the answers are well-defined: A 25% probability to earn "about 0" actually means a 25% probability to earn between 0 and 200.

We will later adopt the interpretation shown in Figure (b), however as illustrated in Figure (c) other interpretations are also possible. Here, the cutoff point between 0 and 400 is at 100. This implies that only amounts up to 100 would be considered "about 0" and the answer reveals a probability of 25% to earn between 0 and 100. More generally, the cutoff point can fall at any point in the interval between these amounts, each cutoff point giving rise to a different interpretation of the data.

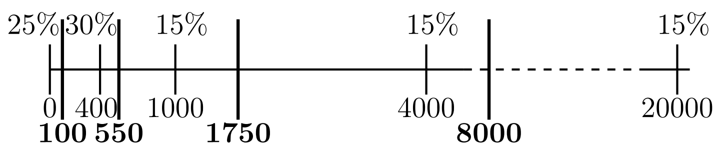
The difficulty in analyzing the data lies in the fact that different interpretations are possible, and that it is impossible to know which is correct. We address this issue in different ways. First, we conduct a simple test of anchoring that does not force us to assume a particular interpretation of the question (Section 4.3.2). Second, to be able to make



(a) Respondent's answers



(b) Possible interpretation



(c) Alternative interpretation

Figure 4.2: Answers and interpretations of the expectation question. Figure (a) shows the respondents' answer to the expectations question. Figures (b) and (c) show two different interpretations of the answers with different cutoff points (in bold) between the amounts.

more general comparisons of the elicited distributions we do impose one particular interpretation (Section 4.3.3). However, in the Appendix we show that the results are robust to varying this assumption and that they also withstand introducing a fair amount of bias against the result.

4.3.2 A simple test of anchoring

In this section, we will do a test of anchoring that does not force us to make an assumption about the interpretation of the expectation question. For this, consider the sets of very low amounts and of high amounts (see Table 4.2). The high amounts are substantially higher than the very low amounts. However, the amounts almost coincide at one point. Under the very low amounts we ask the probabilities to earn 4000 and 20000 (or more), while under the high amounts we ask the probabilities to earn 5000 and 20000.

By comparing the reported probabilities to earn 4000 and 5000 or less when using these different sets of amounts, we will be able to test

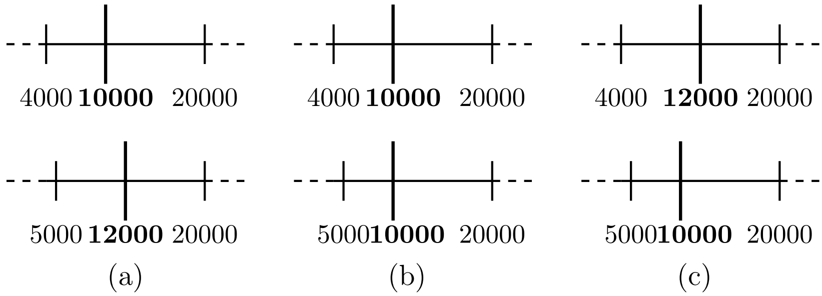


Figure 4.3: Three examples of cutoff points when eliciting expectations at amounts 4000 and 20000 and at 5000 and 20000. The cutoff points in Figures (a) and (b) satisfy the monotonicity assumption, while the cutoff points in Figure (c) do not.

for anchoring. To see why, consider how respondents should report their expectations if they are not affected by anchoring. As discussed before, they should give the probability to earn "about 4000" and "about 5000". Doing so requires setting a cutoff-point between 4000 and 20000 in the case of the very low amounts, and between 5000 and 20000 in the case of the high amounts.

Figure 4.3 gives some examples of how these cutoff points might be set. For instance, in Figure (a) the cutoff point between 4000 and 20000 is 10000 implying that "about 4000" includes amounts up to 10000. The cutoff-point for the between 5000 and 20000 is 12000. So the cutoff point for the interval 5000-20000 is the highest. In Figure (b) the cutoff points are the same for both intervals while in Figure (c) the cutoff point for the interval 4000-20000 is the highest.

Note that the pattern in Figure (c) is odd. While "about 4000" includes any amount up to 12000, "about 5000" only includes amounts up to 10000. Hence, increasing the amount asked about, leads to a decrease in the cut-off considered. We will assume that such situations can not exist by imposing the following *monotonicity* assumption: When eliciting expectations over amounts x, y and over x', y, where $x \geq x'$, the cutoff point between x and y is at least as big as the cutoff point between x' and y. Hence, increasing the amount asked about, while leaving the following amount unchanged, can not decrease the amount at which expectations are reported. This seems like a reasonable assumption and we will impose it in what follows.

If the monotonicity assumption is satisfied, and there is no anchoring, how should the answers to the different questions compare? For this let us compare *cumulative probabilities* at 4000 and 5000. In the case of the very low amounts, the cumulative probability at 4000 is the sum of the probabilities to earn 0, 400, 1000 or 4000. With a cutoff point c between 4000 and 20000, this cumulative probability is the probability to earn at most c. Similarly, for the high amounts the cumulative probability to earn 5000 (the sum of the probabilities to earn 0, 2000 or 5000) is the probability to earn at most c', where c' is the cutoff point between 5000 and 20000. If the monotonicity assumption is satisfied, $c' \geq c$ and, so, the probability to earn at most c' is at least as high as the probability to earn at most c. So, if there is no anchoring, the cumulative probability at 5000 (elicited using the high amounts) should be at least as big as the cumulative probability at 4000 (using the very low amounts).

On the other hand, if there is anchoring we should see the opposite. Indeed, the cumulative probability at 5000 is elicited when using the high amounts while the one at 4000 is elicited using the very low amounts. Anchoring implies that when expectations are elicited over bigger amounts, people would report higher expectations. This means that they should report a lower cumulative probability to earn at most a given amount. Hence, if there is anchoring we expect that the cumulative probability at 5000 using the high amounts to be smaller than the one at 4000 when using the very low amounts.

In Table 4.3(1) we compare the reported cumulative probability at 4000 using the high amounts and at 5000 using the very low amounts. The cumulative probability at 4000 is significantly lower (12 percentage points) than the one at 5000. As discussed, assuming monotonicity these answers are not consistent, but they are compatible with anchoring.

A very similar test can be done at one other place:[7] We elicit expectations at 20000 and 100000 in the set of high amounts and at 25000 and 100000 for the very high amounts. As before, by monotonicity the cumulative probability at 25000 (very high amounts) should be bigger than the cumulative probability at 20000 (high amounts). Table 4.3(2) reports the differences and shows that the cumulative probability when using the bigger amounts is again smaller, although the difference is small and far from significant. This might be because the high and very high amounts are both quite big and lead to smaller framing effects.[8]

[7]Doing this test requires finding two similar intervals. For all other possible combinations of intervals, one of the two amounts of the interval is at least twice the corresponding amount in the other interval. The only exception is 1000-4000 (very low amounts) and 1000-2500 (low amounts). However, while the other comparisons we do are biased against finding a framing effect, this would be biased in favour of it. So, while we do see higher expectations at the bigger set of amounts using this comparison, this is not necessarily a sign of anchoring.

[8]In the next section we compare reported expectations under all four sets of

	(1) Cumulative probability at 4000 or 5000	(2) Cumulative probability at 20000 or 25000	(3) Combination of comparisons (1) and (2)
Bigger amounts	-0.122*	-0.0214	-0.0750**
	(0.0625)	(0.0500)	(0.0314)
Comp. (1): Cum Prob at 4000/5000			-0.124***
			(0.0305)
Constant	0.705***	0.782***	0.804***
	(0.0418)	(0.0369)	(0.0244)
N	64	57	121

Standard errors in parenthesis. In col (3) standard errors are clustered at the level of the respondent. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4.3: A test of anchoring: comparing respondents' reported expectations at a particular point in the distribution.

Nonetheless, note that the test is biased against finding an effect and in the absence of any anchoring we would expect the cumulative probability at 25000 to be bigger. Not seeing this is also somewhat suggestive of anchoring.

Finally, to maximize power, we also combine the two previous tests in Table 4.3 (3). Here, we pool all observations of columns (1) and (2). The dummy "bigger amounts" captures whether the cumulative probability is elicited using the bigger set of amounts of the comparison at hand (high amounts at 5000 and very high amounts at 25000). We also control for which of the two comparisons the observation is coming from. As expected, using the bigger amounts again leads to smaller cumulative probabilities (implying higher expectations), this time significant at the 95% level.

In conclusion, we see that there is a framing effect: respondents report significantly higher expectations when they are elicited over higher amounts. This result does not require a particular assumption on how

amounts (see Figure 4.6). We will see that reported expectations increase systematically as the amounts increase, as one would expect under anchoring. However, differences between two subsequent sets of amounts (such as the high amounts and very high amounts) are not big. Significant differences appear when comparing amounts further apart, such as the very low amounts and the high amounts or the low and the very high amounts.

respondents interpret the questions and, instead, only relies on monotonicity of responses. However, the downside of this analysis is that it compares responses only at one particular point in the distribution, does not exploit all data, and does not allow to quantify the magnitude of the framing effect across the distribution. In the next section, we impose stronger assumptions to be able to make such comparisons.

4.3.3 Comparing elicited distributions

In this section, we will again study whether there is a framing effect: When expectations are elicited over higher amounts, do people report higher expectations? However, we will now do this by estimating the full distributions that are reported and we will compare respondents' answers by comparing these distributions. To do this, we proceed in three steps. First, we impose an assumption on the interpretation of the expectation question (that is, on the position of the cut-off points). Second, we fit distributions that model the reported expectations. Finally, we compare these distributions. Note, however, that while we impose an assumption, the results are robust to alternative assumptions and they also withstand a fair amount of bias against the results (see the Appendix).

The interpretation of the expectation question we will assume is illustrated in Figure 4.2(b): We assume that the respondent assigns the probability to earn a certain amount to the proposed amount that is closest to it. For instance, earning 150 is considered "about 0" and not "about 400", while earning 250 would be considered "about 400". This implies that the cut-off point between two subsequent amounts is exactly in the middle between two subsequent amounts: For the very low amounts, the cut-off point between 0 and 400 is at 200 and the cut-off point between 400 and 1000 is at 700. The same assumption is made for the other sets of amounts. For instance, for the high amounts the first two amounts are 0 and 2000 and, so, the cut-off point is in the middle, at 1000. Table 4.8 shows all cut-off points for all sets of amounts under this assumption.

Having imposed this assumption, we can now interpret the responses as probabilities to earn at most a certain amount. This is illustrated in Figure 4.4. The respondent reports 25% probability to earn "about 0", which implies that he has a probability of 25% to earn at most 200, the cut-off point. Next, he reports 30% probability to earn "about 400". Hence, the cumulative probability to earn at most 700, the next cutoff point, is 55% $(25 + 30)$.

More generally, for every cut-off point we can calculate the probability to earn at most that amount. As illustrated in Figure 4.4, this

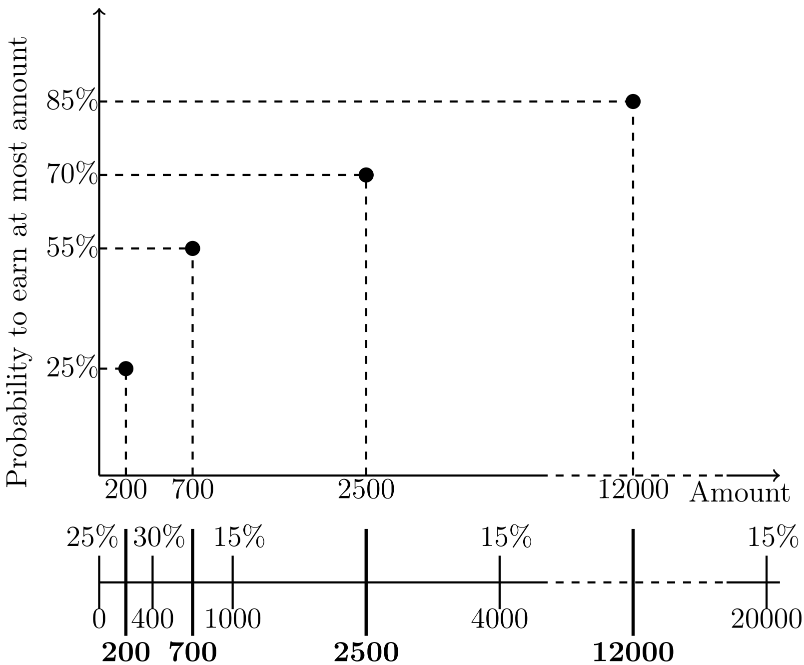


Figure 4.4: Translating answers to the expectations question into points in the cumulative density function.

gives us four points in the cumulative density function (CDF) of the respondent. This, however, does not suffice to compare responses. Indeed, respondents who are presented different sets of amounts have different cutoff points and thus reveal different points in their CDF: When using the lowest amounts, we get the cumulative density at points 200, 700, 2500 and 12000; when using the high amounts, we get the density at points 1000, 3500, 12500 and 60000.

To be able to compare responses across sets of amounts we fit distributions. More precisely, we consider all respondents that are presented a given set of amounts and fit a lognormal distribution reflecting the average distribution for these respondents. To do this, we find for each set of amounts the parameters μ (median) and σ (shape) of the distribution that minimizes the squared errors:

$$\sum_{i}\sum_{p=1}^{4}[\text{Cum-dens}_{i,p} - \text{CDFLOGN}(\text{Amount}_{i,p};\mu;\sigma)]^2 \tag{4.1}$$

where for each of the four points p, Cum-dens$_{i,p}$ is the cumulative density that respondent i associates to the Amount$_{i,p}$ and CDFLOGN is the CDF of the lognormal distribution. Figure 4.5(b) shows the fitted distribution for the very low amounts. The dots show the individual responses, while the crosses give the average of these responses at each amount. As one would expect, the fitted distribution fits the average cumulative densities quite tightly.

The methodology outlined here - fit a lognormal distribution for each set of amounts and compare these distributions - begs several questions. First, why use a lognormal distribution? There are several reasons. First, the lognormal distribution fits the data very well (see Figure 4.6). Second, also the actual returns of those migrants (the empirical counterpart of the fitted distributions, assuming rational expectations) seem to follow a lognormal distribution[9]. Finally, lognormal distributions are often used when studying (skewed) income distributions. Dominitz & Manski (1997), for instance, use it to model measured income expectations in the United States.

The second methodological question is why to fit only one distribution per set of amounts, as opposed to fitting a distribution for every individual. Here, the reason is that some respondents give degenerate answers. For instance, one respondent put all the stones on the first

[9]Diagnostic plots reveal that the lognormal fits these data quite well. The fit is similar or somewhat better than a 2-parameter Weibull and much better than a standard normal distribution

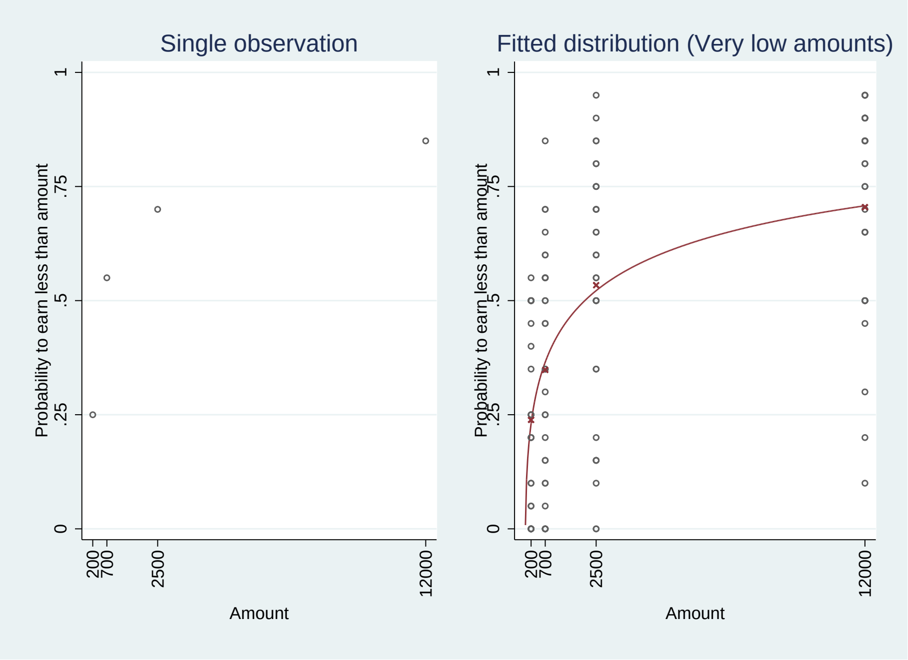


Figure 4.5: Left: Data points given by a single respondent showing the elicited probability to earn at most a given amount. Right: All answers for the group of respondents presented the “very low amounts". Crosses show the averages of the answers and the line is the fitted lognormal distribution for these respondents.

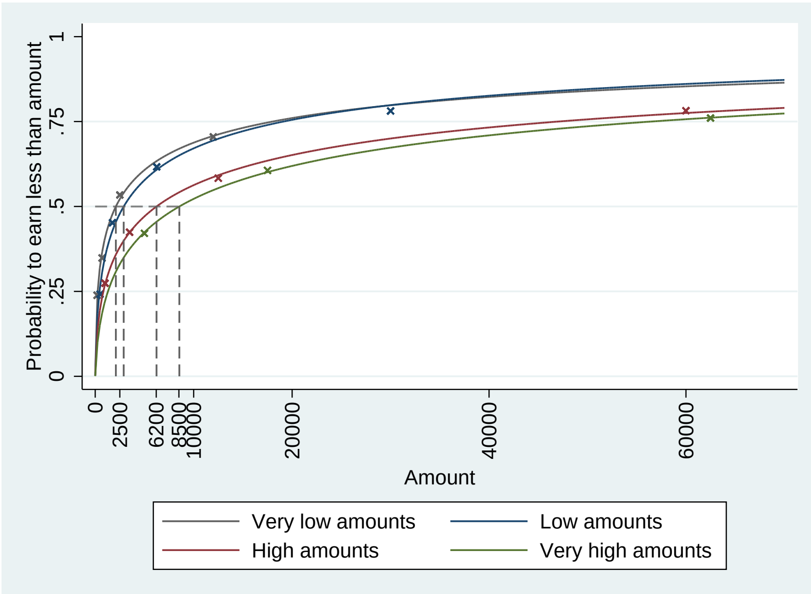


Figure 4.6: Fitted lognormal distributions (CDFs) for the four groups of amounts. The crosses indicate the average responses for each amount

amount. This is not an “irrational" answer. Unsurprisingly, this happened for the set of very high amounts where this answer implies that there is a 100% probability to earn less than 5000. This respondent might simply have very low expectations that are not captured by the amounts proposed to him. However, fitting a distribution just for this respondent is problematic: The distribution with median 0 and shape parameter 0 actually fits the data points perfectly but surely underestimates the true expectations of this person. Dropping the observation is equally problematic: We would systematically drop people with very low expectations from the group of very high amounts (this person would probably not be dropped when using lower amounts which would allow him to properly specify his low expectations). So, in analyzing this person’s data at an individual level we would always need to make an arbitrary assumption about his actual expectations. However, this person’s responses can perfectly be used when “grouping" observations the way we propose to do.

To study the framing effect, we compare the distributions we fitted for each set of amounts. Figure 4.6 shows the four fitted distributions. Additionally, it shows average cumulative densities for each set of amounts.

This graph strongly suggests that there is a framing effect: the higher the amounts presented to the respondents, the higher the expectations.

First, we see that higher amounts lead to a consistent rightward shift of the CDFs, which implies that expectations are higher. The graph illustrates this with the example of the median (i.e., the amount associated to the probability of 50% to earn that amount or less). For the very low and low amounts, the median is about 2000, while it is 6200 and 8500 for the high and very high amounts. The same pattern holds true at other percentiles of the distribution.

At some points in the graph, the same pattern also emerges without having to resort to the fitted distributions. At about the 60th percentile we see that there are true data points for the low, high and very high amounts: The 60th percentile is about 8000 for the low amounts; 13000 for the high amounts; and 18000 for the very high amounts. Again, the higher the amounts, the higher the expectations.

To test whether these differences are statistically significant, we jointly estimate the distributions and compare their parameters. To simplify the exposition we regroup answers in 2 groups: The "lower amounts" (low and very low amounts) and "higher amounts" (high and very high amounts)[10]. The next graph shows the distributions fitted for each of these groups.

Table 4.4 reports the difference between these distributions. We compare the distributions in different ways. First, we look at differences in percentiles (median, first quartile and third quartile) of the distributions. The median, when using the higher amounts is about 3 times higher than when using the lower amounts. All these differences are significant.

The second comparison is through a Kolmogorov-Smirnov test. Here, the test statistic is the maximal vertical distance between the two CDFs. (Note that the differences in percentiles look at *horizontal* differences). In the appendix we explain in more detail how we execute the Kolmogorov-Smirnov test. The difference between the distributions is highly significant.

In choosing the test statistics to compare the distributions, we aim to rely as little as possible on the distributional assumption we make. While we could compare, for instance, the means and standard deviations of the fitted distributions, this comparison would be affected by the exact fit of the extreme tail (for which there is no data available). For this reason, we restrict to direct comparisons of points in the CDF, which

[10]Though increasing power, the result in this paper do not rely on regrouping answers in this way. Significant differences also arise when comparing different subgroups one by one. When comparing subsequent sets of amounts (very low vs low, low vs high and high vs very high) differences are not significant. However, significant differences arise for all other comparisons (very low vs high, very low vs very high and low vs very high) for both the median and the KS-test.

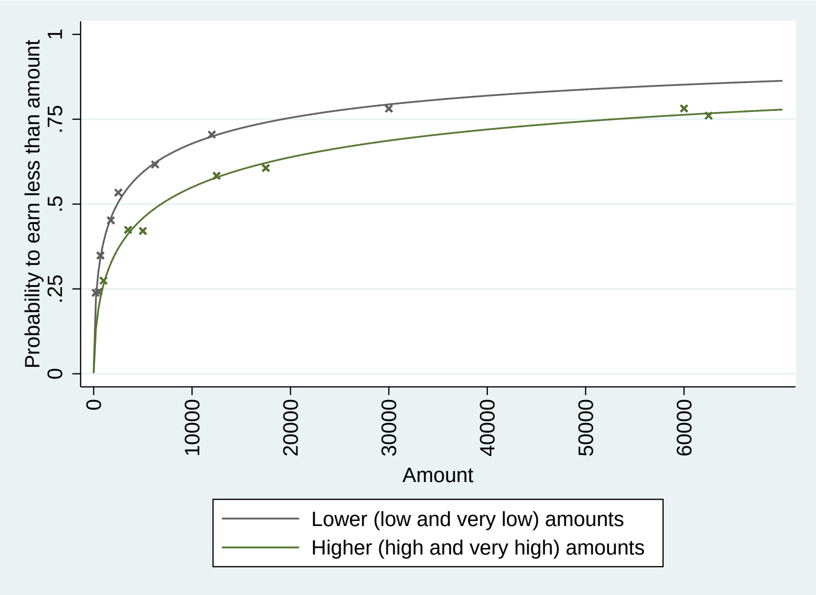


Figure 4.7: Fitted lognormal distributions (CDFs) for the lower and higher amounts. Here, the “low amounts" regroup the low and very low amounts, and the “higher amounts" the high and very high amounts. The crosses indicate the average responses for each amount.

	Lower amounts	Higher amounts	Difference (p-value): Framing effect
Median	2406	6862	0.020
First Quartile	306	903	0.081
Third Quartile	18928	52145	0.020
Kolmogorov-Smirnov test			0.014

Standard errors are clustered at the level of the respondent. The p-value of the Kolmogorov-Smirnov test is based on simulated critical values.

Table 4.4: Comparison of fitted distribution for lower amounts and higher amounts.

		Lower amounts	Higher amounts	Difference (p-value): Framing effect
Has experience mining (N=69)	Median	2198	4047	0.259
	First Quartile	286	462	0.500
	Third Quartile	16888	35486	0.138
	Kolmogorov-Smirnov test			0.276
No experience mining (N=43)	Median	2837	13288	0.037
	First Quartile	344	2166	0.070
	Third Quartile	23418	81541	0.103
	Kolmogorov-Smirnov test			0.020

Standard errors are clustered at the level of the respondent. The p-value of the Kolmogorov-Smirnov test is based on simulated critical values. When comparing the framing effects for respondents with and without experience (the relevant "interaction"), the results are significant for the median and first quartile (p-values of 0.04 and 0.1, resp.) and insignificant for the third quartile (p-value 0.27).

Table 4.5: Comparison of fitted distributions, separating respondents with and without experience mining.

have been directly elicited from the respondents. In the same logic, we restrict all comparisons to places of "common support" in the distribution and do not compare, for instance, responses at the 90th percentile of the distribution.

We next consider heterogeneous effects. In particular, we separate people who have experience in working on the mines (who have worked at least once on a mine) and people who have no experience. Table 4.5 shows the same analysis as before while separating these two groups.

We see that the framing effects are biggest on people who do not have experience in mining. The median is almost 5 times higher (13288 vs 2837) when presented the higher amounts and these differences are significant. For people who have experience mining, the data suggests that there are framing effects, but these effects are much smaller and far from significant. In the appendix we show that these heterogeneous effects are robust to controlling for the respondent's level of education, age and wealth.

The most natural interpretation of these heterogeneous effects is that people who have experience working in the mines are more knowledgeable about the income distribution, and as a consequence are less affected by the framing.

4.4 Interpretation: Anchoring

4.4.1 Anchoring

The most likely explanation for the framing effects we find is anchoring: “the disproportionate influence on decision makers to adjust to an initially presented value" (Furnham & Boo 2011). This phenomenon is, according to Kahneman (2011), “one of the most reliable and robust results in experimental psychology".

In a classical experiment, Tversky & Kahneman (1974) found, for instance, that a number that results from spinning a wheel strongly influences answers to the subsequent question: “What is the percentage of African countries in the United Nations?". In another study, Englich et al. (2006) find that experienced judges give substantially higher fictitious prison terms when the roll of a die comes up 6 instead of 1.

In our case, respondents might anchor on the amounts we presented to them and report higher expectations when they are presented with higher amounts. This interpretation is particularly appealing since we find that people least knowledgeable about the distribution are most affected - and least knowledgeable people are typically most affected by anchoring (Furnham & Boo 2011; Wilson et al. 1996; Northcraft & Neale 1987; Luchini & Watson 2013).

However, from this we should not conclude that anchoring is not an issue when measuring expectations among people who know the distribution well. Anchoring is typically also found among knowledgeable people, only less so. Suggestive examples are real-estate agents on housing prices (Northcraft & Neale 1987) or, as mentioned before, experienced judges on prison terms (Englich et al. 2006).

More generally, the anchoring effect has been extensively studied in psychology and provides important insights in the phenomenon. (See Furnham & Boo (2011) for a recent literature review and Kahneman (2011) for an excellent layman's introduction.) For instance, it is commonly believed that people with little knowledge try to extract information from the anchor, and display anchoring effects for this reason. However, it has been shown that anchoring effects are equally strong when the anchor is deliberatively uninformative and people are told that they will suffer from anchoring problems. Anchoring effects thus arise for other reasons. In Section 5 we will come back to this literature to understand when anchoring effects might arise and what can be done about them.

4.4.2 Alternative explanation: Noise

An alternative explanation that needs to be addressed is noise: The possibility that the results are generated by people failing to understand the expectations questions. What would happen in our case if some respondents would fail to understand the questions and would randomly spread out the stones (probability masses) across different amounts? Respondents presented high amounts would put probability mass on these amounts and would seem to have high expectations. This would be coherent with the basic framing effect we find.

It is clear that the answers we get are not “pure noise", i.e., a completely random allocation of stones. Answers are influenced by the amounts that are proposed. For instance, when the highest amount is 20 000 (or more) respondents allocate a probability of about 30 %; when it is 500 000 (or more), they allocate about 10 %. Nonetheless, this does not suffice to rule out the noise explanation: If part of the respondents would give useful answers, but others random answers, we would see such patterns in the data, as well as framing effects.

It is however also unlikely that the results are generated by such “partial noise". First, the framing effect is not correlated with levels of education - and one would expect uneducated farmers to have more problems understanding the questions. Likewise, one might worry that the people who claimed that they might leave for the mines, but did not end up doing so, were not replying to the questions seriously. However, also this variable is not correlated with the framing effect[11].

Most importantly, the framing effect is correlated with having experience, exactly as we would expect if anchoring is at play.[12] All in all, anchoring seems to be a much more likely explanation than noise for the pattern of results we see.[13]

[11] Concretely, we redid the analysis of heteregenous effects (Table 4.4) using the variables “Has formal education", “Can read and write" and “Actually left" and looked at the differences in framing effects (the “interaction") for the different parameters. The smallest p-value across all these specications is 0.27 and almost all p-values are well above 0.5

[12] Additionally, restricting the sample to only the respondents interviewed by the most diligent enumerators (as judged by the author) yields essentially the same results.

[13] Another way to assess the quality of the data is to assess the accuracy of the measured expectations, but this is fraught with difficulties in our case. The framing effects of inexperienced respondents is so big that it is impossible to say what their expectations are, so we should restrict to experienced respondents. Additionally, we observe returns over an entire year (which can involve a few days up to several months of work) and have to compare it to expectations over a month of work. Finally, outcomes depend on the price of gold and the quality of the sites which

4.5 Do anchoring effects really cause problems?

So far we have argued that anchoring effects can substantially affect reported expectations. But does this really matter? Will anchoring lead to problems when analyzing expectations data? In short, our answer is yes. In particular the fact that anchoring effects can be heterogeneous can create problems in many types of analysis.

The answer this question more precisely, let us distinguish three typical uses of expectations data: 1) To measure the level of expectations per se, 2) as an explanatory variable of interest and 3) as a control.

First, anchoring effects are most problematic when trying to measure the actual level of expectations. Measuring the level of expectations would allow to tell whether rational expectations hold or whether expectations are too optimistic or pessimistic (see, e.g., McKenzie et al. (2013)). Clearly, data that suffers from important anchoring effects - which can arbitrarily increase or decrease expectations - can not be used to answer such a question.

Second, expectations can be the explanatory variable of interest, for instance, when studying whether expectations about returns to migration affect the decision to migrate. While the measured expectations could serve as a proxy for "true" expectations, problems arise when anchoring effects are heterogeneous. In our case, less experienced miners have bigger anchoring effects and so the measured expectations capture both experience and expectations, and results should be interpreted accordingly. A correlation between measured expectations and the decision to migrate could very well be driven by the fact that more experienced miners are more likely to migrate.

Finally, the measured expectations might be used to control for confounding factors in a regression. This is probably the least problematic case, but we should still be aware of the fact that the measured expectations capture other factors when anchoring effects are heterogeneous and the measured expectations are only an imperfect proxy. As such, they can help to attenuate, but not eliminate, any bias that would arise from omitting expectations from the regression.

In the last two cases, where the problem is that measured expectations capture other variables, problems could be solved by controlling

fluctuate yearly and are strongly correlated among all respondents. Nonetheless, comparing expectations of experienced migrants to historical realizations suggests that they overestimate the probabilities of earning very low and very high amounts and underestimate probabilities to get amounts in between. This corresponds quite well to the way mining is perceived: An all or nothing activity. Nonetheless, all this tells us little about the quality of the data.

for these variables. For instance, in a regression studying the decision to migrate, we could control for experience in migration. The key here is that we should control for factors that correlate with the anchoring effect and explain the outcome of interest, even if they do not correlate with the "true" expectations themselves. Anchoring effects typically correlate strongly with knowledge about the answers and it is well worth trying to control for such factors (e.g., for migration this could be having experience, knowing people with experience or having a network at the destination). When such variables are not available, one can also ask respondents if they are "confident" about their answer. This has been shown to correlate with the anchoring effect (Chapman & Johnson 1994), and controlling for it could alleviate the bias.

Anchoring effects can thus pose problems, even when only a proxy for expectations is needed. While some of these problems appear relatively innocent, it is important to keep in mind that anchoring effects can be very big, as they are for inexperienced respondents in this study[14]. In this case, the variation induced by anchoring can be much larger than the variation in expectations itself and these problems should not be ignored.

4.6 Factors (not) contributing to anchoring

Since anchoring effects cause problems when analyzing expectations data, there are two key questions. First, in which situations can we expect anchoring effects? Second, can we make methodological choices that avoid (or reduce) anchoring effects?

To answer these questions, we will address in this section different factors that contribute to anchoring effects.

Using a self-anchored support. The anchoring effect arose in this experiment because respondents anchored to the amounts proposed to them. A common alternative is to avoid proposing amounts by asking the respondent's expectations of his minimum and maximum income. Based on these amounts one can then come up with a set of amounts in between this minimum and maximum and elicit expectations at these amounts.

Some authors believe that using such self-anchored support allows to reduce anchoring problems. Morgan & Small (1992) argue that amounts proposed by the interviewer might be considered "objectively reason-

[14] Kahneman (2011) argues that the "anchoring effect" is typically around 55%. This implies that increasing the anchor by 100 increases the answer by 55.

able" and so influence the respondent. However, anchoring happens equally well (and equally strongly) for self-generated or obviously non-informative anchors (Furnham & Boo 2011). Hence, this does not seem to be a reason why anchoring might be reduced using this methodology.[15]

Nonetheless, this methodology might still reduce anchoring for other reasons. Delavande et al. (2011a) use it and show convincingly that which number is asked for first (the minimum or the maximum) does not influence answers. Hence, people do not seem to anchor to the first amount they have to come up with, suggesting that there might not be any anchoring problems at all. Intriguingly, however, they find that standard deviations are higher when using this methodology. This could be compatible with people anchoring on both amounts (the minimum and the maximum) and overestimating the probability of extreme events. So, using self-anchored supports might not remove all anchoring effects, but it is a promising method to reduce them as much as possible.

Choices when using a pre-determined support. When using the method of pre-determined support, as we do, there are still methodological issues to resolve. One is the number of amounts (or intervals) to use. We used a relatively small number of amounts (5) and it is possible to use more intervals. Additionally, how amounts are spread out over the entire support also matters. Should they be equally spaced or organized differently?

To assess the impact of such choices, it's useful to consider the mechanism through which anchoring works. A leading explanation is that people use an “anchoring and adjustment heuristic" which states that the anchor is used as an initial answer which is then moved towards the range of plausible answers (Tversky & Kahneman 1974). The bias arises because one stops adjusting “too early" upon reaching a first reasonable value. For instance, in estimating the number of african countries when the anchor is 200, one would first evaluate 200 as a possible answer. One would then start to decrease it steadily until reaching the first value that

[15] We can also test for this using our data. If anchoring occurs because people believe there is information in the amounts presented to them, then people should update their income expectations based on these amounts. People presented the higher expectations should then have higher expectations and be more likely to actually leave for the mines the following year. However, we do not observe this. In fact, if anything, people presented higher amounts are *less* likely to leave to the mines. Though the results are estimated too imprecisely to draw strong conclusions, this “wrong" sign does allow us to reject what needed to be tested: We can reject (at the 95% level) that being presented the higher amounts increases by 2 percentage points or more the probability to leave to the mines. So, it does not seem that people have actually updated their expectations by being presented the higher amounts.

looks reasonable.

In our context, spreading probability mass equally over all amounts first, and then adjusting away from this answer would be one way to implement this heuristic. If this is true, it would be mainly the distribution of amounts that would influence answers. For instance, using multiple high amounts (instead of just one) would lead to an initial answer that puts weight on each of these amounts and thus much weight on high amounts. Adjusting away from these high expectations, but stopping too early, would still lead to reporting high expectations. If this is correct, the answers could be biased towards the true distribution by picking for instance the deciles of the true distribution as the amounts.

Note, however, that to set up experiments in this way, researchers need to know the true distribution before conducting the experiment, which requires investing resources. While knowing the true distribution does not in itself solve the anchoring the problem, it might nonetheless allow to set up the expectation question in such a way as to minimize biases because of anchoring and to know the likely direction of the biases.

As is clear from the discussion, we are far from understanding how different methods induce (or not) anchoring. While the psychological literature can tell us a lot on when anchoring might be relevant, it typically deals with anchoring on a single number, rather than the multiple amounts needed to elicit expectations. While some first steps have been taken (e.g., Delavande et al. 2011a show that it does not matter which amount is elicited first), there is definitely much scope for methodological research on which methods least induce anchoring. In this respect, experiments that compare self-anchored supports with different ways of using pre-determined supports would be particularly interesting.

Context of the interview Anchoring effects, while robust, do vary with circumstances. Respondents who are less tired or in a better mood (Bodenhausen et al. 2000), seem to be less affected by anchoring. This suggests that if expectations are being measured as part of a long questionnaire, these questions should come early on in the questionnaire.

Unfortunately, some other potential strategies to mitigate anchoring effects do not seem to work. Warning people upfront about the risks of anchoring or given them incentives to answer correctly, do not seem to cause big decreases in the anchoring effect (Tversky & Kahneman 1974; Furnham & Boo 2011).

Knowledge about the distribution. As mentioned before, respondents who are less knowledgeable about the returns have bigger anchoring effects. While the researcher can, of course, not affect this parameter, it does allow to assess whether anchoring problems are likely to be big.

In this paper, respondents had personally seen very few realizations of the distribution: Some never went mining and the others only observed a realization once per year. Moreover, the form of the distribution also matters. Here the distribution was strongly skewed with few people earning big amounts. Undoubtedly, it is more difficult to assess expectations over such a distribution than over a simple symmetric distribution.

In this sense, it is perhaps unsurprising that we find big anchoring effects. Likewise, Delavande et al. (2011a) work with boat owners who see daily realizations of their symmetric income distribution and are very knowledgeable about the distribution. Accordingly, they do not find anchoring effects. An interesting question is how big anchoring effects are in more intermediate cases.

4.7 Conclusion

We conducted an experiment on eliciting expectations and find strong anchoring effects: Expected median incomes increase up to 4-fold when higher amounts are used to elicit expectations. This effect is mainly concentrated on people with no experience in mining, who are least likely to be knowledgeable about the income distribution.

These results suggest that caution is needed when using expectations data elicited from people with little knowledge about the issue, even when the data is to be used as a simple proxy for expectations.

However, more research is needed on determining when anchoring effects arise and what can be done to reduce them. The current setting, where the income distribution is skewed and with much uncertainty about this distribution, is particularly challenging. Anchoring effects might be smaller in other settings. Additionally, other elicitation methods might suffer less from anchoring. To shed light on this, experiments comparing self-anchored supports with different ways of using pre-determined supports would be particularly interesting.

Appendix: Robustness tests

Adding controls

We redo the regressions on heterogeneous effects while adding controls. While the main causal effect (the framing effect) should be well-identified, the heterogeneous effect with respect to having experience is not. If, for instance, educated people are less affected by the framing and educated people have less experience in mining, we would see the same pattern of results.

To control for such confounding factors, we redo the analysis by subgroup based on the controls. For instance, to control for the ability to read and write, we separately do the analysis for people who can read and write and for those who can't. We then also take the averages of these estimations. These averages are the closest in spirit to a normal regression with controls for confounding factors and we will focus on these averages.

Table 4.6 reports the results controlling for the ability to read and write, having formal education, wealth and age. Here, the difference between the ability to read and write and having formal education lies in following alphabetization courses. The wealth variable is a consumption proxy based on the Progress out of Poverty Index (PPI), which combines different wealth proxies.

Despite the limited numbers of observations, the heterogeneous results with respect to experience hold up very well when controlling for these variables. Inside each subgroup we consider, the framing effect is bigger for people without experience than it is for people with experience.

Moreover, the "average" framing is never significant for people with experience and consistently significant for respondents without experience. The only exception to the latter is when separating respondents based on formal education. This is unsurprising given the very small number of respondents without experience and with formal education (10). For the ones without formal education and without experience the framing effect is significant and even for the 10 respondents without formal education the framing effect is big (2448 vs 19794). When using the more balanced (and thus more powerful) measure for education "can read and write", the average framing effect for respondents without experience is significant again. The heterogeneous effects thus seem to be robust to controlling for levels of education.

Varying assumptions on interpretation of the question

The results in Section 4.3.3 rely on a particular assumption of how respondents interpret the expectations question. Recall that respondents are asked about the probability to earn "about 0", "about 400" and so on, and they need to settle on a particular interpretation to say what "about 0" and "about 400" exactly mean.

In Section 4.3.3 we have assumed that the respondent assigns the probability to earn a certain amount to the proposed amount that is closest to it. For instance, earning 150 is considered "about 0" and not "about 400", while earning 250 would be considered "about 400". This implies that the cut-off points between the amounts are exactly in the middle between two subsequent amounts: For the very low amounts, the cut-off point between 0 and 400 is at 200 and the cut-off point between 400 and 1000 is at 700.

To test the robustness of our results to this assumption, we redo the analysis imposing different assumptions. Our assumption was that the cutoff falls in the middle of the interval between two subsequent amounts. Our alternative assumptions will be that the cutoff point is at other places in the interval. That is, we will still assume that the same "rule" is applied to every interval, but we will assume different rules. Concretely, we assume that the cutoff point falls either (a) completely to the left of the interval (b) at 25% (c) in the middle (d) at 75% and (e) to the right of the interval. If we take the example of the amounts 0 and 400, this implies that the cutoff points fall at (a) 0, (b) 100, (c) 200, (d) 300 and (e) 400. Note that the assumption we imposed in Section 4.3.3 is assumption (c) here, i.e., that the cutoff is in the middle of the interval.

In Figure 4.8 and Table 4.7 we report the results of imposing these different assumptions. These figures and tables are identical to Figure 4.7 and Table 4.4 in the body of the paper, except for the alternative assumption. We can see that moving the cutoff points to the left (a and b) reduces expectations while moving them to the right increases expectations (d and e).

However, the anchoring effect is highly robust to varying these assumptions. Under all assumptions, reported expectations are substantially higher when elicited over the higher amounts. Moreover, these differences are significant under all assumptions and at all places in the distribution, except for the first quartile under assumptions (a) and (b). However, even in these cases the reported expectations under the higher amounts are almost 3 times higher (539 against 175 and 520 against 189) and these differences are almost significant. Overall, the results seem robust to these alternative assumptions.

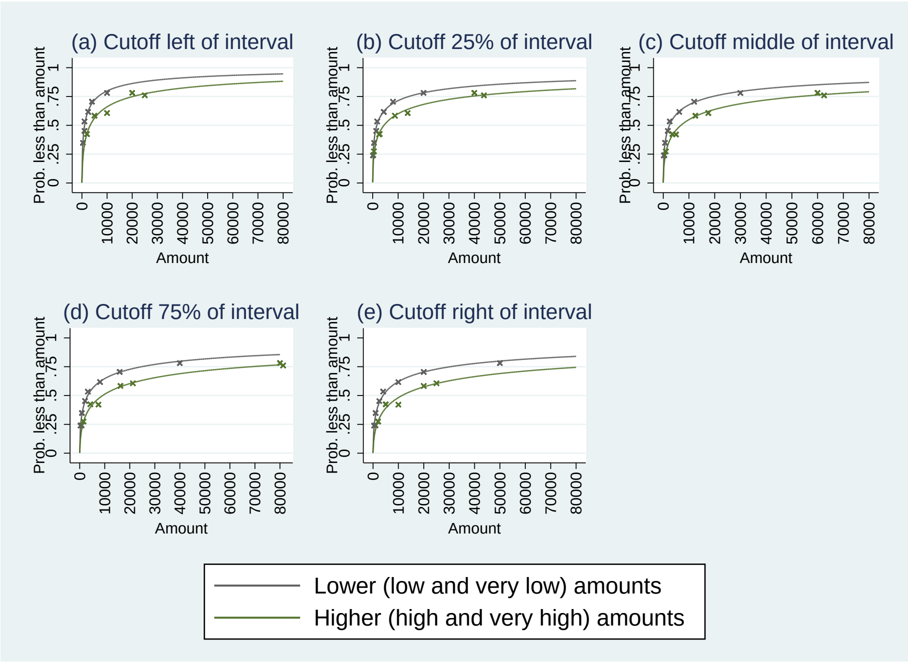


Figure 4.8: Comparison of reported expectations using different assumptions on how respondents interpreted the expectations quesion. Figure (c) replicates Figure 4.7. The other figures impose alternative assumptions. Table 4.7 tests whether these differences are significant.

Introducing bias against the results

In the previous section, we used different assumptions on how respondents interpret the expectation question. Moreover, we imposed the same assumption on all respondents, no matter which set of amounts they were presented. While this makes intuitive sense, different respondents are presented different amounts and so do not necessarily need to interpret question in the same way.

For this reason, we will test now if the results are robust to imposing different assumptions for different respondents. In particular, we will do this in such a way that biases the results against our finding that higher amounts lead to higher expectations. A measure of the robustness of our results is then how much of this bias these results can withstand.

To bias the results against the finding we proceed as follows. For the lower amounts, we impose assumptions that increase expectations. For the higher amounts we do the opposite and decrease expectations. To do this, note that level of expectations depends on the place of the cutoff point in the interval. Moving the cutoff point completely to the left of the interval (assumption (a) above) gives the lowest possible expectations compatible with the answers. On the other hand, moving the cutoff completely to the right gives the highest possible expectations.

To bias the results against us, we start again from the central assumption that the cutoff point is in the middle of the interval, and this for all respondents. Next, we simultaneously move the cutoff points to the left for respondents given the higher amounts and to the right for the ones given the lower amounts. The extreme case is where we assume the cutoff point is completely to the right for the higher amounts and completely to the left for the lower amounts. We will say that this is a 100% bias against our results. However, there are also intermediate cases. For instance, assuming that the cutoff point is at 25% of the interval for the lower amounts and at 75% of the interval for the higher amounts is a 50% bias against our results. More precisely, for a level of bias $b \in [0, 100]$, the cutoff point c_L between two amounts x and y when using the lower amounts is

$$c_L = \frac{x+y}{2} + \frac{b}{200}(y-x)$$

while for the higher amounts the cutoff point c_H is

$$c_H = \frac{x+y}{2} - \frac{b}{200}(y-x)\,.$$

The results of systematically increasing the amount of bias against our results is shown in Figure 4.9. Introducing more and more bias systematically increases expectations for the lower amounts and decreases it for the higher amounts.

How much bias can the results withstand? Up to 25% bias the difference remains statistically significant at several points in the distribution (median and 3th quartile). After this, the difference becomes insignificant but remains there up to 75% of bias, at which point the two distributions coincide. At the maximal level of bias the pattern is inversed and respondents presented higher amounts would report somewhat lower expectations.

Overall it thus seems that the results withstand a fair amount of bias and that one needs to introduce quite a bit of bias to remove it completely. As an example, consider again the intervals used in Section 4.3.2: 4000-20000 for the very low amounts and 5000-20000 for the high amounts. One would expect respondents to report expectations at a

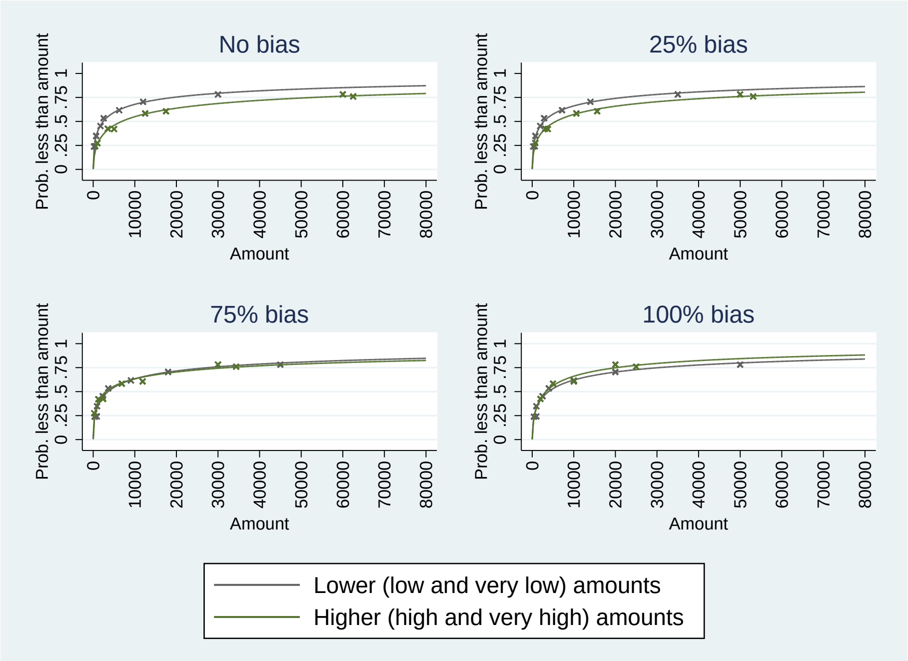


Figure 4.9: Comparison of reported expectations when introducing increasing levels of bias against the results. Up to 25% of bias, the results remain significant. To completely eliminate the differences in reported expectations, we need to introduce 75% bias against the results.

quite similar point in the distribution. However, when imposing 75% bias we would assume that he reports expectations at 18000 for the very low amounts and at 6875 for the high amounts. This is a very strong violation of the monotonicity assumption we presented in that section. This thus seems to be a very strong bias.[16] We thus need quite a big bias against the results to completely remove the anchoring effect, which suggests that the results are quite robust.

Methodology for Kolmogorov-Smirnov test

One of the tests we use in this paper is a Kolmogorov-Smirnov (K-S) test (Kolmogorov 1933; Smirnov 1933). Given that we could not apply the test in the simplest way, we lay out here the choices made in implementing this test.

[16]For a 25% bias, the cutoff point would be 14000 for the very low amounts and at 10625 for the high amounts, which is a smaller but still substantial violation of monotonicity.

The test-statistics of the K-S test, denoted k, to compare two distributions with CDFs $\widehat{F}_1$ and $\widehat{F}_2$ is the maximal distance between the two CDFs:

$$k = \max\{\widehat{F}_1(x) - \widehat{F}_2(x) \mid x \in R\}$$

In a normal application this test statistic needs to be compared to the critical values supplied by Smirnov (1933) in his original article to determine a p-value and significance of this test.

We could not apply this procedure for two reasons. First, we do not have data over the entire support of the distribution. In particular, we did not ask respondents about the probability of earning extremely big payouts. It would thus be inappropriate to compare fitted distributions at points for which no data was collected. Second, the critical values provided by Smirnov (1933) are based on a DGP in which all data points were generated independently from the CDF. In our case we get 4 data points from each respondent, and these points are clearly not independent. We can thus not use the usual critical values.

To solve the first problem, we will restrict the analysis to the part of the support over which we elicited expectations: between 1000 and 30000[17]. The test statistic will thus be

$$k = \max\{\widehat{F}_1(x) - \widehat{F}_2(x) \mid 1000 \leq x \leq 30000\}$$

To solve the second problem, we will simulate the critical values. To avoid having to guess an appropriate DGP for our problem, we simulate these critical values in a non-parametric way using a bootstrap-like procedure. Concretely, we follow the following procedure to do the kolmogorov-smirnov test. For each individual i, we have four points $k = 1, 2, 3, 4$ in his CDF, where $y_{i,k}$ denotes the probability to earn at most the amount $x_{i,k}$.

To calculate the test-statistic k:

1. Fit a log-normal distribution $\widehat{F}_H$ using the sample $(y_{i,k}, x_{i,k})$ for the sample of individual presented high amounts

2. Fit a log-normal distribution $\widehat{F}_L$ using the sample $(y_{i,k}, x_{i,k})$ for the sample of individual presented low amounts

3. The test-statistic is: $k = \max\{\text{abs}(\widehat{F}_H(x) - \widehat{F}_L(x)) \mid 1000 \leq x \leq 30000\}$

[17] 1000 is the lowest amount presented used for the respondents presented the higher amounts; 30000 is the highest for the ones presented the lower amounts

To simulate the critical values and calculate an associated p-value:

1. Fit a log-normal distribution $\widehat{F}_0$ using all observations. Under the null of no differences between F_H and F_L, the distribution $\widehat{F}_0$ is a consistent estimate of these distributions.

2. Calculate the predicted values: $\widehat{y}_{i,k} = \widehat{F}_0(x_{i,k})$

3. Calculate the residuals: $\widehat{\epsilon}_{i,k} = \widehat{y}_{i,k} - y_{i,k}$

4. Repeat the following 1000 times:

   (a) Draw an independent random variable u_i with a Rademacher two-point distribution: $u_i = 1$ with probability 0.5 and $u_i = -1$ with probability 0.5.

   (b) Define the bootstrap sample $y^*_{i,k} = \widehat{y}_{i,k} + u_i \cdot \widehat{\epsilon}_{i,k}$

   (c) Fit a lognormal distribution $\widehat{F}^*_H$ using the bootstrap sample $(y^*_{i,k}, x_{i,k})$ for the sample of individual presented high amounts.

   (d) Fit a lognormal distribution $\widehat{F}^*_L$ using the bootstrap sample $(y^*_{i,k}, x_{i,k})$ for the sample of individual presented low amounts. The bootstrapped test-statistic is: $k^* = \max\{\text{abs}(\widehat{F}^*_H(x) - \widehat{F}^*_L(x)) \mid 1000 \leq x \leq 30000\}$.

5. The p-value associated to the kolmogorv-smirnov test is the proportion of bootstrapped test-statics k^* that is bigger than the test statistic k.

			Lower amounts	Higher amounts	Difference (p-value): Framing effect
Has experience mining	Can read and write (N=18)	Median	2554	3463	0.788
		K-S test			0.566
	Can not read and write (N=51)	Median	2043	4129	0.257
		K-S test			0.310
	Average of above	Median	2298	3796	0.436
No experience mining	Can read and write (N=18)	Median	2210	13431	0.236
		K-S test			0.170
	Can not read and write (N=25)	Median	3265	13084	0.084
		K-S test			0.052
	Average of above	Median	2738	13257	0.017
Has experience mining	Has formal education (N=12)	Median	1927	1376	0.712
		K-S test			0.914
	Has no formal education (N=57)	Median	2280	4613	0.242
		K-S test			0.306
	Average of above	Median	2103	2995	0.448
No experience mining	Has formal education (N=10)	Median	2448	19794	0.569
		K-S test			0.390
	Has no formal education (N=33)	Median	3010	12616	0.040
		K-S test			0.032
	Average of above	Median	2729	16205	0.333
Has experience mining	Is wealthier than median (N=30)	Median	1864	7124	0.175
		K-S test			0.132
	Is less wealthy than median (N=39)	Median	2485	2854	0.815
		K-S test			0.938
	Average of above	Median	2175	4989	0.177
No experience mining	Is wealthier than median (N=24)	Median	2988	13657	0.226
		K-S test			0.140
	Is less wealthy than median (N=19)	Median	2649	12792	0.027
		K-S test			0.074
	Average of above	Median	2818	13224	0.007
Has experience mining	Is older than median (N=26)	Median	1952	488	0.274
		K-S test			0.214
	Is less old than median (N=43)	Median	2312	9472	0.019
		K-S test			0.000
	Average of above	Median	2132	4980	0.109
No experience mining	Is older than median (N=26)	Median	2370	5418	0.272
		K-S test			0.364
	Is less old than median (N=17)	Median	4941	40856	0.070
		K-S test			0.036
	Average of above	Median	3655	23137	0.034

Standard errors are clustered at the level of the respondent. The p-value of the Kolmogorov-Smirnov test is based on simulated critical values. The line "average of above" contains the average of the preceding two lines. The p-value on this line thus refers to the average of the differences in medians between the two frames, i.e., it measures the significance of the average framing effect.

Table 4.6: Comparison of fitted distributions, separating respondents with and without experience mining and separating respondents by level of education, age and wealth.

	Lower amounts	Higher amounts	Difference (p-value)
Median	1059	3280	0.020
First Quartile	175	539	0.157
Third Quartile	6406	19958	0.004
Kolmogorov-Smirnov test			0.018

(a) Cutoff points to the left of the interval

	Lower amounts	Higher amounts	Difference (p-value)
Median	1624	4425	0.033
First Quartile	189	520	0.113
Third Quartile	13976	37672	0.026
Kolmogorov-Smirnov test			0.014

(b) Cutoff points at 25% of interval

	Lower amounts	Higher amounts	Difference (p-value)
Median	2406	6862	0.020
First Quartile	306	903	0.081
Third Quartile	18928	52145	0.020
Kolmogorov-Smirnov test			0.004

(c) Cutoff points in the middle of the interval (central assumption)

	Lower amounts	Higher amounts	Difference (p-value)
Median	3173	9167	0.017
First Quartile	413	1245	0.073
Third Quartile	24387	67479	0.019
Kolmogorov-Smirnov test			0.014

(d) Cutoff points at 75% of interval

	Lower amounts	Higher amounts	Difference (p-value)
Median	3934	11426	0.016
First Quartile	516	1571	0.070
Third Quartile	30020	83122	0.019
Kolmogorov-Smirnov test			0.010

Standard errors are clustered at the level of the respondent. The p-value of the Kolmogorov-Smirnov test is based on simulated critical values.

(e) Cutoff points at the right of the interval

Table 4.7: Replication of Table 4.4, but with alternative assumptions on how respondents interpreted the expectations quesion.

Category	Am. 1	Cutoff 1	Am. 2	Cutoff 2	Am. 3	Cutoff 3	Am. 4	Cutoff 4	Am. 5
Very low amounts	0	200	400	700	1000	2500	4000	12000	20000
Low amounts	0	500	1000	1750	2500	6250	10000	30000	50000
High amounts	0	1000	2000	3500	5000	12500	20000	60000	100000
Very high amounts	0	5000	10000	17500	25000	62500	100000	300000	500000

Table 4.8: Amounts used to elicit expectations and cutoff points in the middle between subsequent amounts.

Chapter 5

Time efficient algorithms for robust estimators of location, scale, symmetry and tail heaviness[1]

[1] This chapter is co-authored with Vincenzo Verardi and Catherine Vermandele.

Abstract

The analysis of the empirical distribution of univariate data often includes the computation of some location, scale, skewness and tails heaviness measures which are estimates of specific parameters of the underlying population distribution. Several measures are available but they differ in terms of Gaussian efficiency, robustness with respect to outliers, and meaning in case of asymmetric distributions. We first briefly compare, for each type of parameter (location, scale, skewness, and tail heaviness), the "classical" estimator based on (centered) moments of the empirical distribution, an estimator based on specific quantiles of the distribution, and an estimator built on the basis of pairwise comparisons of the observations. This last one always performs better than the other estimators, namely in terms of robustness, but requires at first sight a heavy computation time of an order of n^2. Fortunately, as explained in Croux & Rousseeuw (1992), the algorithm of Johnson & Mizoguchi (1978) allows to substantially reduce the computation time to an order of $n \log n$ and, hence, allows to use the robust estimators based on pairwise comparisons even in very large datasets. This has motivated us to program this algorithm and make it available in Stata: we describe in this paper the sketch of the algorithm and the associated Stata commands. Finally, we illustrate on real data the interest of the computation of these robust estimators by involving them in a normality test of the Jarque-Bera form (Jarque & Bera (1980); Brys et al. (2004b)).

5.1 Introduction

When analyzing univariate data, a key task is to estimate location, scale, skewness and tails heaviness parameters of the underlying distribution. These will together provide a very good characterization of this distribution. Several estimators for these parameters are available but they do not all share the same properties; they differ in terms of Gaussian efficiency, robustness with respect to outliers, smoothness of the influence function and meaning in case of asymmetric distributions.

In this paper, we will systematically compare, for each type of parameter (location, scale, skewness and kurtosis), three estimators of different natures. The first one, generally considered as the "classical" estimator, is based on the first, second, third or fourth (centered) moment of the empirical distribution; the second one is defined on the basis of specific quantiles of the distribution; the third one is built on the basis of pairwise comparisons of the observations. They will be compared in terms of breakdown point (i.e., maximal outlier contamination they withstand), Gaussian efficiency (i.e., relative asymptotic variances) and smoothness of the influence function (i.e., relative sensitivity of the estimator to changing a fraction of points in the sample).

As will be explained in the paper, the estimators of the third category perform very nicely, both in terms of efficiency and robustness. This contrasts with the other estimators. The classical estimators of the first category are highly efficient, but not robust to outliers. The quantile-based estimators of the second category have the opposite property: they are very robust, but not efficient. The pairwise-based estimators of the third category, however, are typically as robust as the quantile-based ones, but more efficient (though not always as efficient as the classical estimators). In this sense, these pairwise-based estimators combine the best of two worlds.

However, because these estimators are based on pairwise comparisons, it is often thought that the heaviness of their computation makes them unfeasible in practice. To overcome this apparent excessive computational complexity, we follow the idea already developed in Croux & Rousseeuw (1992) which consists in applying the very efficient deterministic algorithm of Johnson & Mizoguchi (1978) that allows to reduce the computation time from an order of n^2 to an order of $n \log n$. Stata commands are programmed to make the estimators of the third type available for applied researchers.

The paper is structured as follows. In Section 5.2, we introduce various estimators for the location, scale, skewness and tails heaviness of the distribution from which the data have been generated and compare their (asymptotic) Gaussian efficiency and robustness properties. In Section 5.3, we show how these estimators may be used in practice to test the normality of the distribution, following the idea of the Jarque & Bera (1980) statistical test even when outliers are present. This motivates the presentation, in Sections 5.4 and 5.5, of the sketch of an efficient algorithm and of the associated Stata commands. Section 5.6 is devoted to an example and Section 5.7 concludes.

5.2 Estimators of location, scale, skewness and heaviness of the tails

Many measures of location, scale, skewness and kurtosis or heaviness of the tails have been proposed and studied in the statistical literature. This section is devoted to the comparison of the (asymptotic) Gaussian efficiency and robustness performance of three different types of estimators: (i) the "classical" estimators, based on (centered) moments of the distribution; (ii) the estimators built from specific quantiles of the distribution; (iii) the estimators defined on the basis of pairwise comparisons or combinations of the observations.

5.2.1 Definitions

The different estimators will be compared in terms of breakdown value, Gaussian efficiency and influence functions.

The *asymptotic breakdown value* is the maximal outlier contamination an estimator can withstand before breaking down, that is, leading to arbitrary values.

Gaussian efficiency is related to the asymptotic variance of the estimator under a Gaussian distribution. The lower the asymptotic variance, the more efficient the estimator.

The *influence function* at a point x measures the effect on the estimator of a perturbation of the distribution by adding a small probability mass at point x. We are mostly interested in whether the influence function is *bounded* and *smooth*. When it is unbounded, the effect of an outlier on the estimator can be arbitrarily large. This implies that the estimator is not robust to outliers. When the influence function is smooth, a small change in a data point only has a small effect on the estimator. For this reason, smoothness of the influence function tends to improve efficiency.

Finally, note that the influence function can also be used to obtain asymptotic confidence intervals (see Hampel et al. (1986), p. 85 and 226).

	Type	ASV$(\cdot, \Phi)$	Asymptotic breakdown val.	Bounded IF?
$\bar{x}$	Classical	1	0%	NO
$Q_{0.5}$	Quantile-based	$\pi/2$	50%	YES
HL	Pairwise-based	$\pi/3$	29%	YES

Table 5.1: A comparison of the three location estimators' performance with respect to Gaussian efficiency, asymptotic breakdown value, and boundedness of the influence function. HL is less robust but more efficient than its robust alternative, the median.

Alternatively, jackknife can also be used to obtain confidence intervals.

5.2.2 Location estimators

There is apparently a consensus in applied statistics about the fact that the sample mean ($\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$) and the sample median ($Q_{0.5} = F_n^{-1}(0.5)$) are two complementary location estimators: the mean is very efficient in case of Gaussian data but fragile to outliers (and meaningless in case of highly asymmetric data) while the median is very robust (and meaningful in case of asymmetries) but rather inefficient. Both are extensively used in practice.

Less well-known is the midpoint estimator based on pairwise comparisons introduced by Hodges & Lehmann (1963). The *Hodges-Lehmann estimator* is defined by $\mathrm{HL} = \mathrm{med}\left\{\frac{x_i+x_j}{2}; i < j\right\}$.

In terms of robustness, the HL, like the median, outperforms the mean. Figure 5.1 shows the influence functions under the standard Gaussian distribution Φ. The influence function of the HL and the median, in contrast to the mean, are bounded. As a consequence they both have positive asymptotic breakdown values (see Table 5.1). The median, with a breakdown value of 50%, is more robust to outliers than HL.

In terms of efficiency, the mean is known to be the most efficient location estimator. Among HL and the median, HL comes out as the most efficient one, with an asymptotic variance under the standard Gaussian distribution of $\pi/3$ against $\pi/2$ for the median (see Table 5.1). This is also illustrated by the influence functions, where the influence function of the HL appears as a smooth version of that of the median.

Although the HL has nice properties, being based on pairwise comparisons, it seems to have a high computational complexity and not to be usable in big samples. However, as will be explained in Section 5.4, a simple algorithm proposed by Johnson & Mizoguchi (1978) al-

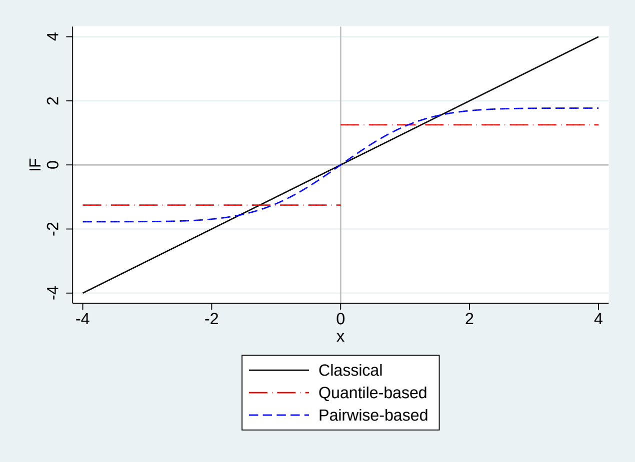


Figure 5.1: Influence functions of the location estimators under the standard Gaussian distribution. The influence function of HL is bounded and appears as a smooth version of that of the median.

lows to substantially reduce this computational complexity from $O(n^2)$ to $O(n \log n)$. We programmed this estimator in Stata and describe the associated command in Section 5.5. We nevertheless do not believe that it brings enormous advantages with respect to the median and therefore do not strongly advice to use it systematically instead of the median.

5.2.3 Scale estimators

To estimate the scale parameter of the underlying distribution, the classical estimator is the *standard deviation*: $s = \left[\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2\right]^{1/2}$. It is the most efficient estimator of the scale parameter σ in case of Gaussian data. But, like the mean, it completely lacks robustness: its influence function is unbounded (see Figure 5.2) and, as a consequence, it has an asymptotic breakdown value of 0% (e.g., Rousseeuw & Croux (1993), p. 1275).

There are however several robust alternatives to the standard deviation. First, there are two alternatives based on quantiles. A commonly used one is the *interquartile range*: $\text{IQR} = d \times (Q_{0.75} - Q_{0.25})$ where setting $d = 0.7413$ ensures consistency for σ at Gaussian distributions. A second one is the $\text{MAD} = b \times \text{med}_i|x_i - \text{med}_j x_j|$ where $b = 1.4826$ makes it consistent for σ at Gaussian distributions. As discussed below, the MAD is very robust, but has the downside that it aims at *symmetric* distributions only: it essentially finds the symmetric interval (around the median) that contains 50% of the data, which does not seem to be a nat-

	Type	ASV$(\cdot, \Phi)$	Asymptotic breakdown val.	Bounded IF?
s	Classical	0.5	0%	NO
IQR	Quantile-based	1.3605	25%	YES
Qn	Pairwise-based	0.6077	50%	YES

Table 5.2: A comparison of three scale estimators' performance with respect to Gaussian efficiency, asymptotic breakdown value, and boundedness of the influence function. Qn is more robust and more efficient than its robust alternative, IQR.

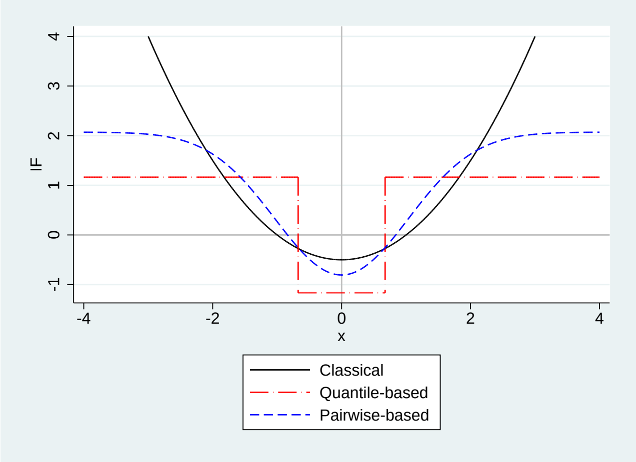


Figure 5.2: Influence functions of the scale estimators under the standard Gaussian distribution. The influence function of Qn is bounded and appears as a smooth version of that of IQR.

ural approach at asymmetric distributions. The interquartile range does not have this problem, as the quartiles need not be equally far away from the median. Because the MAD has this problem of assumed symmetry, we focus mainly on IQR from now on.

Finally, a very interesting but relatively unknown scale estimator is the Qn statistic of Rousseeuw & Croux (1993): $\text{Qn} = d \times \{|x_i - x_j|; i < j\}_{(k)}$, where $d = 2.2219$ ensures consistency for Gaussian distributions and $k = \binom{h}{2} \cong \binom{n}{2}/4$ with $h = [n/2] + 1$. In other words, the statistic Qn corresponds approximately to the 25th percentile of the $\binom{n}{2}$ distances $\{|x_i - x_j|, i < j\}$.

The Qn estimator generally outperfoms the other robust estimators, both in terms of efficiency and robustness. First, its Gaussian efficiency, at 83%, is surprisingly high and substantially higher than that of IQR and MAD (see Table 5.2).

Also in terms of robustness, Qn outperforms the other estimators. It has an asymptotic breakdown value of 50% which is higher than that of IQR. Finally, Qn is also applicable to assymetric distributions and its influence is smooth, in contrast to that of IQR (see Figure 5.2).

Despite its very nice statistical performances, it seems at first sight difficult to use Qn in practice because of the high computational complexity needed to compute it. Indeed, according to its definition, we have to determine an order statistic of $\binom{n}{2}$ pairwise differences. However, as for the midpoint estimate the algorithm proposed in Johnson & Mizoguchi (1978) can be used. We programmed the Qn estimator in Stata with this efficient algorithm and called the command `sqn`. A detailed description of this command is available in Section 5.5.

5.2.4 Skewness estimators

Skewness is often measured by the Fisher estimator: $\gamma_1 = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{s}\right)^3$. Since this skewness measure relies on the mean and the standard deviation, it is not surprising that its resistance to outliers is null. More precisely, its asymptotic breakdown value is equal to 0% and its influence function is unbounded (see Figure 5.3 and, for example, Groeneveld (1991)). Alternative estimators of skewness, such as $\frac{\bar{x} - \text{mode}}{s}$ and $\frac{\bar{x} - Q_{0.5}}{s}$ proposed by Pearson, still rely on the standard deviation and are thus just as fragile with respect to outliers as the classical skewness estimator.

Fortunately, there again exist several robust alternatives. First, Hinkley (1975) proposed the quantile-based estimator

$$\text{SK}_p = \frac{(Q_{1-p} - Q_{0.5}) - (Q_{0.5} - Q_p)}{Q_{1-p} - Q_p} = \frac{Q_p + Q_{1-p} - 2Q_{0.5}}{Q_{1-p} - Q_p},$$

	Type	ASV$(\cdot, \Phi)$	Asymptotic breakdown val.	Bounded IF?
Fisher	Classical	6	0%	NO
$\mathrm{SK}_{0.25}$	Quantile-based	1.8421	25%	YES
MC	Pairwise-based	1.25	25%	YES

Table 5.3: A comparison of three skewness estimators' performance with respect to Gaussian efficiency, asymptotic breakdown value, and boundedness of the influence function. The Medcouple is as robust and more efficient than its robust alternative, $\mathrm{SK}_{0.25}$.

where $0 < p < 0.5$. This is in fact a generalization of Yule and Kendall's skewness estimator which can be obtained by setting p equal to 0.25: $\mathrm{SK}_{\mathrm{YK}} = \mathrm{SK}_{0.25}$.

An alternative robust skewness operator, called *medcouple*, has been proposed by Brys et al. (2004a). It is based on pairwise comparisons and replaces the quantiles Q_p and Q_{1-p} in SK_p by actual data points. More precisely, let $x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(n)}$ denote the n order statistics associated to the sample, then

$$\mathrm{MC} = \mathrm{med}_{x_{(i)} \leq Q_{0.5} \leq x_{(j)}} h\left(x_{(i)}, x_{(j)}\right)$$

where, for all $x_{(i)} \neq x_{(j)}$, the kernel function h is given by

$$h\left(x_{(i)}, x_{(j)}\right) = \frac{\left(x_{(j)} - Q_{0.5}\right) - \left(Q_{0.5} - x_{(i)}\right)}{x_{(j)} - x_{(i)}} \ .$$

For the special case $x_{(i)} = x_{(j)} = Q_{0.5}$, we define the kernel as follows: let $m_1 < \ldots < m_k$ denote the indices of the order statistics that are tied to the median $Q_{0.5}$ (that is $x_{(m_l)} = Q_{0.5}$ for all $l = 1, \ldots, k$). Then,

$$h\left(x_{(m_i)}, x_{(m_j)}\right) = \begin{cases} -1 & \text{if } i + j < k + 1 \\ 0 & \text{if } i + j = k + 1 \\ +1 & \text{if } i + j > k + 1. \end{cases}$$

Because of the denominator, it is clear that $h\left(x_{(i)}, x_{(j)}\right)$, and hence MC, always lies between -1 and $+1$ (like SK_p). The medcouple is 0 for symmetric distributions while it is positive and negative for respectively right and left tailed distribution.

Overall, the medcouple outperforms Yule and Kendall's skewness estimator. In terms of robustness they are comparable: they both have an

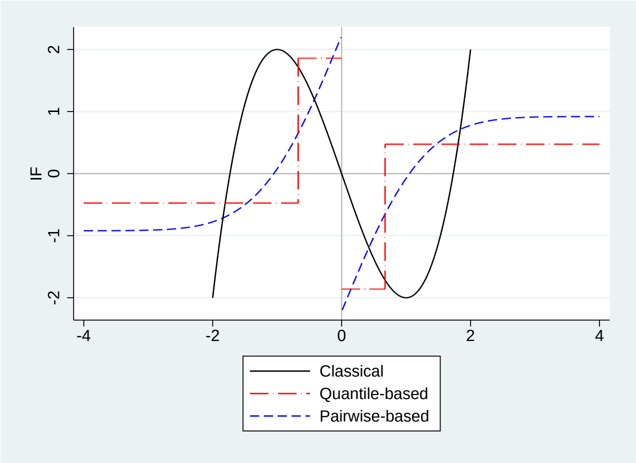


Figure 5.3: Influence functions of the skewness estimators under the standard Gaussian distribution. The influence function of the Medcouple is bounded and appears as a smooth version of that of $\text{SK}_{0.25}$

asymptotic breakdown value of 25%[2] (see Table 5.3) and have bounded influence functions. The influence function of the medcouple is however smoother than that of SK_p.

The big difference between the medcouple and SK_{YK} lies in efficiency. The Gaussian efficiency of the former is substantially higher than that of SK_{YK}. In fact, the Gaussian efficiency of these robust estimators is even better than that of the classical Fisher estimator of skewness.

As for Qn, at first sight, the computational complexity of MC is of the order of $O(n^2)$. As for HL and Qn, the Johnson and Mizoguchi algorithm can be used to compute MC with a complexity of $O(n \log n)$. We programmed this skewness estimator in Stata and describe the commands in Section 5.5.

5.2.5 Tails heaviness estimators

Tail heaviness is often measured using kurtosis: $\gamma_2 = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{x_i - \mu_n}{\sigma_n} \right)^4$. The parameter γ_2 is equal to three for distributions with tails similar to the normal, bigger than three for leptokurtic distributions (i.e., with heavier tails than the normal) and smaller than three for platokurtic distributions (i.e., with lighter tails than the normal). However, this parameter also measures the peakedness of a distribution and it is difficult

[2]The asymptotic breakdown value of SK_p is $100p$% and is thus bigger than 25% when setting p higher than 0.25 (the value used in SK_{YK}). Doing this however also reduces efficiency.

	Type	ASV$(\cdot, \Phi)$	Asymptotic breakdown val.	Bounded IF?
γ_2	Classical	24	0%	NO
$\mathrm{LQW}_{0.25}$ ($\mathrm{RQW}_{0.25}$)	Quantile-based	3.71	12.5%	YES
LMC (RMC)	Pairwise-based	2.62	12.5%	YES

Table 5.4: A comparison of the tail heaviness estimators' performance with respect to Gaussian efficiency, asymptotic breakdown value, and boundedness of the influence function.

to grasp what kurtosis really estimates. Another disadvantage of the kurtosis is that its interpretation and consequently its use is restricted to symmetric distributions. Moreover, as usual for estimators relying on the mean and the standard deviation, the kurtosis coefficient is very sensitive to outliers in the data (0% asymptotic breakdown value and unbouded influence function[3]; see Figure 5.4 and Ruppert (1987))

To overcome these problems, Brys et al. (2006) have proposed two measures of *left* and *right* tail weight for univariate continuous distributions. These measures have the advantage that they can be applied to symmetric as well as asymmetric distributions that do not need to have finite moments; moreover, they unambiguously measure tail heaviness (not peakedness) and they are robust against outlying values.

More precisely, Brys, Hubert and Struyf defined *left* and *right* tail measures as measures of skewness that are applied to the half of the probability mass lying to the left, respectively to the right, side of the median $Q_{0.5}$. As measures of skewness they use the two robust estimators presented in the previous section: SK_p $(0 < p < 0.5)$ and MC.

Concretely, applying the medcouple to each side of the distribution we get the *Left Medcouple* and the *Right Medcouple*: $\mathrm{LMC} = -\mathrm{MC}(x < Q_{0.5})$ and $\mathrm{RMC} = \mathrm{MC}(x > Q_{0.5})$. Similarly, using the $\mathrm{SK_p}$ skewness estimator leads to the *Left Quantile Weight* and the *Right Quantile Weight*: $\mathrm{LQW}_p = -SK_{p/2}(x < Q_{0.5})$ and $\mathrm{RQW}_p = SK_{p/2}(x > Q_{0.5})$. Thus, $\mathrm{LQW}_{0.25}$, for instance, considers the quantiles $Q_{0.125}$ and $Q_{0.375}$ around the center of the left side of the distribution $(Q_{0.25})$. For these estimators, a higher value of these estimators means a fatter tail. For comparison, note that the tail weights of the normal distribution are 0.2.

The performance of these robust measures of tail heaviness is strictly connected to the performance of their underlying estimators (Medcouple

[3]The form of this influence function shows that contamination at the center has far less influence than that in the extreme tails; this suggests that γ_2 is primarily a measure of tail behavior, and only to a lesser extent of peakedness.

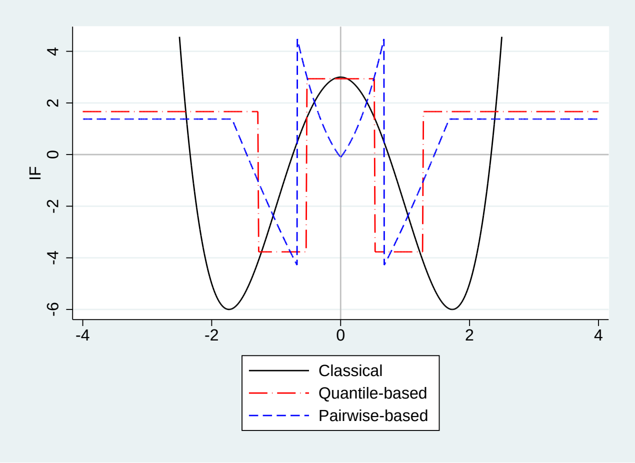


Figure 5.4: Influence functions of the tail heaviness estimators under the standard Gaussian distribution

and $\mathrm{SK_p}$) and so follows the same pattern as in the previous section. The Left and Right Medcouple have a higher Gaussian efficiency than $\mathrm{LQW}_{0.25}$ and $\mathrm{RQW}_{0.25}$ (they are also more efficient than the classical kurtosis for Gaussian data). In terms of robustness, all their influence functions are bounded (see Figure 5.4 and Brys et al. (2006) (p. 740–741)) and the asymptotic breakdown value is the same for Left and Right Medcouple and LQW_p and RQW_p when $p = 0.25$. For the latter estimators increasing p increases robustness, but decreases efficiency; decreasing p does the opposite. This, in fact, points at another advantage of the Left and Right Medcouple: it does not require to (somewhat arbitrarily) fix a value for p.

To conclude, note that the efficient algorithm of Johnson and Mizoguchi allows once again to substantially reduce the computational complexity needed to compute LMC and RMC. These tail weight measures have been implemented as separate options in the medcouple code in Stata (see Section 5.5).

5.3 Normality test based on skewness and tails heaviness estimators

As stated above, these descriptive statistics can be used to characterize the underlying distribution. In particular, they can be used to test for normality. For example, the Jarque & Bera (1980) test relies on the classical skewness and kurtosis coefficients to test for normality. More precisely, under the normality assumption ($\gamma_1 = 0$ and $\gamma_2 = 3$), we can write

$$\sqrt{n}\begin{pmatrix} \gamma_1 \\ \gamma_2 - 3 \end{pmatrix} \xrightarrow{d} \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 6 & 0 \\ 0 & 24 \end{pmatrix}\right),$$

which leads to the Jarque-Bera test statistic:

$$T = n\left(\frac{\gamma_1^2}{6} + \frac{(\gamma_2 - 3)^2}{24}\right) \approx \chi_2^2.$$

The Jarque-Bera test is a very popular and interesting test for normality: it has been shown that, for a wide range of alternative distributions, it outperforms such tests as the Kolmogorov-Smirnov test, the Cramér-von Mises test and the Durbin test. Unfortunately, despite its good power properties and computational simplicity, the Jarque-Bera test is highly sensitive to outliers because it is constructed from the moment-based skewness and kurtosis measures.

Robust alternatives to the Jarque-Bera test have been proposed and studied in Brys et al. (2004b). These authors start from the fact that the Jarque-Bera test can be seen as a special case of the following general testing procedure. Let $\widehat{\boldsymbol{\theta}} = (\widehat{\theta}_1, \ldots, \widehat{\theta}_k)'$ be a vector of estimators of $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)'$ (a vector of characteristic parameters of the underlying distribution) such that, under the null hypothesis of normality,

$$\sqrt{n}\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right)' \xrightarrow{d} \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega});$$

then, the general test consists in rejecting, at level α, the null hypothesis of normality if

$$T = n\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right)' \boldsymbol{\Omega}^{-1}\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right) > \chi^2_{k;1-\alpha},$$

where $\chi^2_{k;1-\alpha}$ is the $(1-\alpha)$-quantile of the chi-square distribution with k degrees of freedom. Brys et al. (2004b) then propose to use, in this general testing procedure, the robust skewness estimator MC and/or the tails heaviness estimators LMC and RMC.

Three tests have been studied. The first one is only based on the skewness estimator MC (the medcouple): in this case, $k = 1$, $\widehat{\theta} = \mathrm{MC}$ and $\Omega = 1.25$. The second one is based on the left and right tail heaviness estimators LMC (left medcouple) and RMC (right medcouple): in this case, $k = 2$, $\widehat{\boldsymbol{\theta}} = (\mathrm{LMC}, \mathrm{RMC})'$, $\boldsymbol{\theta} = (0.199, 0.199)'$ and

$$\boldsymbol{\Omega} = \begin{pmatrix} 2.62 & -0.0123 \\ -0.0123 & 2.62 \end{pmatrix}.$$

The third test combines MC, LMC and RMC: in this case, $k = 3$, $\widehat{\boldsymbol{\theta}} = (\mathrm{MC}, \mathrm{LMC}, \mathrm{RMC})'$, $\boldsymbol{\theta} = (0, 0.199, 0.199)'$ and

$$\boldsymbol{\Omega} = \begin{pmatrix} 1.25 & 0.323 & -0.323 \\ 0.323 & 2.62 & -0.0123 \\ -0.323 & -0.0123 & 2.62 \end{pmatrix}.$$

This last test seems to have the best overall performance.

5.4 Efficient algorithm

The key to efficiently implementing the different estimators based on pairwise combinations described above, is a simple algorithm by Johnson & Mizoguchi (1978). Given a number k and an $n \times q$ matrix M with sorted (non-increasing) rows, this algorithm finds the kth maximal element of M in time $O(n \log n)$. In this section, we give a sketch of this algorithm.

The algorithm proceeds by repeatedly guessing a new candidate for the kth maximum of M. After every such guess, it discards some of the elements of M (because they can not be the kth maximum). In this way, it systematically reduces this set of candidates until it guesses the kth maximum of M. The key to the efficiency of the algorithm is that, at every guess, it (efficiently) discards many elements of M. In this way, the algorithm only needs few attempts before finding the kth maximum.

The algorithm is sketched in somewhat more detail below.

```
1 mat excludeleft = J(n,1,1) // length-n vector with value 1 everywhere
2 mat excluderight = J(n,1,q) // length-n vector with value q everywhere
3 while kth maximum is not found {
4   scalar m = new guess for kth maximum using non-excluded elements of M
5   scalar nrbigger = number of elements a in M with a > m
6   scalar nrsmaller = number of elements a in M with a < m
7   if nrbigger >= k // kth maximum is bigger than m  7
8     excluderight[i] = pos. smallest element bigger than m in row i
9   elseif nrsmaller >= (n*q)-k // kth maximum is smaller than m
10    excludeleft[i] = pos. biggest element smaller than m in row i
11  else // kth maximum equals m
12    m is the kth maximum element of M
13 }
```

In Lines 1 and 2 the algorithm initializes the data structure which maintains which elements have already been discarded. At any time, for any row i, all elements to the left of position excludeleft[i] are discarded. E.g., if excludeleft[1] is 4, the first three elements in row 1 are discarded.

Similarly, excluderight[i] contains the position in row i to the right of which all elements have been discarded.

The loop on Line 3 continues until the kth maximum has been found. It first makes a new guess m for the kth maximum (Line 4) and then calculates the number of elements in M which are bigger and smaller than M (Lines 5 and 6). This can be done efficiently because each of the rows is sorted. Then, if there are more than k elements which are bigger than m (Line 7), we know that the kth maximum must be bigger than m. We can thus discard all elements smaller or equal than m (Line 8). Similarly, we can discard all elements bigger or equal than m if there are more than $(n * q) - k$ elements strictly smaller than m. Finally, if the kth maximum is neither among the elements strictly bigger nor those strictly smaller than m (Line 11), m must be the kth maximum and we have found the kth maximum.

This algorithm requires time $O(n \log n)$ (Johnson & Mizoguchi (1978)). Note, however, that one should not explicitly calculate the entire matrix M (which would be of complexity $O(n^2)$). Indeed the algorithm only needs to inspect some of its elements, and only those elements should be calculated. This substantially reduces the running time of the algorithm. To illustrate the time saving this algorithm achieves, we show in Table 5.5 a comparison of the running time[4] of the different estimators based on pairwise comparisons when using this algorithm, and when using standard algorithms.

5.5 Commands in Stata

We programmed the following commands in Stata to estimate the described statistics using the efficient algorithm of Johnson & Mizoguchi (1978).

- For the HL statistic of Hodges & Lehmann (1963), the command is "`mhl` *varname* [if] [in]".
- For the Qn statistic of Rousseeuw & Croux (1993), the command is "`sqn` *varname* [if] [in]".
- For the medcouple, the command is "`medcouple` *varname* [if] [in]" and three options are available. The first is `lmc` which asks Stata to calculate the medcouple only using observations smaller than the median. This gives a measure of the flatness of the tail on the

[4]We used Stata/SE 12 and a computer with a 2.66 GHz Intel dual-core processor and 2GB of RAM.

n	HL efficient	HL naive	Qn efficient	Qn naive	Medcouple efficient	Medcouple naive
500	0.2	0.1	0.2	0.1	0.1	0.1
1000	0.4	0.5	0.4	0.6	0.3	0.5
2000	0.9	2.9	0.9	2.6	0.7	2.0
5000	2	23	3	22	2	15
10000	6	113	7	117	5	72
50000	46	/	44	/	33	/
100000	105	/	108	/	74	/

Table 5.5: Running time (in seconds) of the implemented estimators (HL, Qn and Medcouple) using the efficient Johnson and Mizoguchi algorithm (left) and using standard algorithms (right). When n is bigger than 10000, the running time of the standard algorithm is not reported as it took too long to compute.

left of the distribution. The `rmc` option does the same for the right tail by focusing on observations larger than the median. Finally, the `nomc` option asks Stata not to calculate the medcouple.

- For the robust test of normality we created the command "`robjb` *varname* [if] [in]". This command implements the test considering both skewness and heaviness of tails by default. Two mutually exclusive options are available: `skewness` and `kurtosis`. If the former is used, a test based exclusively on the skewness is performed while if the latter is called, a test based exclusively on the heaviness of the tails is performed.

5.6 Example

To illustrate the usefulness of the estimators, we will analyze the body weight of 64 different animal species. The dataset we use is available online and have been made available by Rice University, University of Houston Clear Lake and Tufts University.

To start the analysis, we first calculate the classical, quantiles based and pairwise based estimates of location, scale, skewness and heaviness of the tails (see Table 5.6). The classical and quantiles based estimates can easily be calculated using the formulas provided in the theoretical section and using the standard `summarize` and `centile` Stata commands (see the do-file relative to the example for further details). To compute the pairwise based estimates, we have to use the commands we programmed as follows:

```
mhl body
sqn body
medcouple body, lmc rmc
```

	Classical	Quantiles based	Pairwise based
Location	$\bar{x}$: 3 111 355.5	$Q_{0.5}$: 3 500	HL : 94 307.5
Scale	s : 13 033 900	IQR : 166 221.28	Qn : 6 665.438
Skewness	γ_1 : 5.461	$\mathrm{SK}_{0.25}$: 0.976	MC : 0.985
Tails	γ_2 : 32.770	LQW : -0.052 RQW : 0.883	LMC : -0.090 RMC : 0.915

Table 5.6: Classical, based on quantiles, and based on pairwise combinations estimates of location, scale, skewness and tails heaviness.

If we would only look at the classical estimators, we would conclude that the average animal weight is very high but with a huge dispersion. The asymmetry is large and positive and tails are very heavy. When we look at the equivalent robust statistics, we see that the median weight is much lower than the mean weight. The robust dispersion is also much smaller than that suggested by classical estimators and right skewness is extreme. As far as the heaviness of the tails is concerned, the right tail is extremely heavy while the left one is slightly lighter than the left tail of the normal (recall that the normal has tail weights of 0.2 and that higher values indicate heavier tails). When looking at the difference between classical and robust estimators, it is evident that outliers are present in the dataset.

A first thing that we could do to tackle this problem is to transform the data to reduce the excessive importance of very big animals (such as dinosaurs). Given that weights are strictly positive, we consider a logarithmic transformation and re-do the above descriptive statistics analysis (see Figure 5.5 and Table 5.7):

```
gen lbody = ln(body)
mhl lbody
sqn lbody
medcouple lbody, lmc rmc
```

When we do this transformation, we see that the differences between classical and robust estimators become much smaller. Indeed the mean

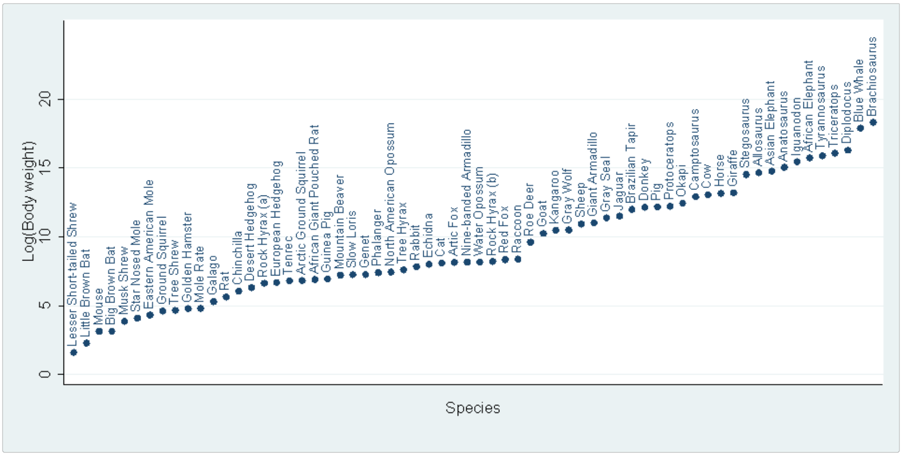


Figure 5.5: Logarithm of the body weights of 64 animal species, in increasing order

	Classical	Quantiles based	Pairwise based
Location	$\bar{x}$: 9.313	$Q_{0.5}$: 8.161	HL : 9.289
Scale	s : 4.135	IQR : 4.207	Qn : 4.281
Skewness	γ_1 : 0.304	$\text{SK}_{0.25}$: 0.465	MC : 0.386
Tails	γ_2 : 2.192	LQW : 0.499 RQW : 0.241	LMC : 0.515 RMC : 0.241

Table 5.7: Classical, based on quantiles, and based on pairwise combinations estimates of location, scale, skewness and tails heaviness for transformed data.

is only slightly larger than the median, the dispersion estimate is very similar for all the methods and the skewness estimate only points towards evidence of moderate positive skewness. As far as the heaviness of the tails is concerned, the classical estimator is close to 3 which is the value of the kurtosis for the normal and therefore points towards standard tails. Nevertheless when we look at the robust estimator for the latter, there is evidence of a heavy left tail. Indeed, the left tail weights of about 0.5 are substantially above 0.2, the tail weight of normal tails. This last point is very important.

Indeed, let us imagine that we want to test for the normality of the log(body) variable. The classical and the robust estimators lead to different results. The standard Jarque-Bera statistic is 2.726 which is much smaller than the critical value of $\chi^2_{2;0.95} = 5.99$. This implies that the standard Jarque-Bera test would not reject the null hypothesis of normality. On the other hand, we can also calculate the robust test statistic involving MC, LMC and RMC:

```
robjb lbody
```

We see that this robust test statistic is equal to 9.266 which is larger than the critical value of $\chi^2_{3;0.95} = 7.815$. This would therefore point towards the rejection of the null hypothesis of normality. This means that even though the logarithmic transformation substantially reduced the effect of atypical individuals, outliers still bias the classical estimations. In particular, we believe that the heaviness of the left tail is not satisfactorily identified by the classical kurtosis coefficient, and this affects the result of the normality test.

5.7 Conclusion

Different statistics are available to estimate the location, the scale, the skewness and the tails heaviness of a distribution. Some of these estimators are based on pairwise comparisons of the observations; these estimators, apparently much more heavy and complex to compute, perform better, namely in terms of robustness and efficiency. The algorithm of Johnson & Mizoguchi (1978) allows to substantially reduce the computation time of these robust estimators from an order n^2 to an order of $n \log n$; this makes it possible to determine these estimators even in very large datasets. In order to make these interesting estimators available for applied researchers, we have programmed them in Stata, following the efficient algorithm of Johnson and Mizoguchi. We have also programmed a robust version of the Jarque-Bera test of normality, for which the test statistic involves some of these estimators of skewness and tails heaviness.

Bibliography

Ahlin, C., & Townsend, R. M. (2007). Using repayment data to test across models of joint liability lending. *The Economic Journal*, *117*(517), F11–F51.

Akerlof, G. A. (1970). The market for "lemons": Quality uncertainty and the market mechanism. *The quarterly journal of economics*, *84*, 488–500.

Andreoni, J., & Sprenger, C. (2012). Risk preferences are not time preferences. *The American Economic Review*, *102*(7), 3357–3376.

Andreoni, J., Sprenger, C., et al. (2010). Certain and uncertain utility: The allais paradox and five decision theory phenomena. Levine's Working Paper Archive.

Arnott, R., & Stiglitz, J. E. (1991). Moral hazard and nonmarket institutions: Dysfunctional crowding out of peer monitoring? *The American Economic Review*, *81*(1), 179–190.

Baksi, S., & Bose, P. (2007). Credence goods, efficient labelling policies, and regulatory enforcement. *Environmental and Resource Economics*, *37*(2), 411–430.

Baland, J.-M., & Duprez, C. (2009). Are labels effective against child labor? *Journal of Public Economics*, *93*(11), 1125–1130.

Banerjee, A., Duflo, E., & Hornbeck, R. (2014). Bundling Health Insurance and Microfinance in India: There Cannot be Adverse Selection if There is No Demand. *The American economic review*, *104*(5), 291.

Basaza, R., Criel, B., & Van der Stuyft, P. (2008). Community health insurance in Uganda: why does enrolment remain low? A view from beneath. *Health Policy*, *87*(2), 172–184.

Basu, A. K., Chau, N. H., & Grote, U. (2006). Guaranteed manufactured without child labor: the economics of consumer boycotts, social labeling and trade sanctions. *Review of Development Economics*, *10*(3), 466–491.

Basu, A. K., & Hicks, R. L. (2008). Label performance and the willingness to pay for Fair Trade coffee: a cross-national perspective. *International Journal of Consumer Studies*, *32*(5), 470–478.

Baumann, F., Oschinski, M., & Staehler, N. (2012). On the effects of fair trade on the welfare of the poor. *Journal of International Development*, *24*(S1), S159–S172.

Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, reciprocity, and social history. *Games and economic behavior*, *10*(1), 122–142.

Besley, T., & Coate, S. (1995). Group lending, repayment incentives and social collateral. *Journal of development economics*, *46*(1), 1–18.

Bhagwati, J. (1995). Trade liberalisation and "fair trade" demands: addressing the environmental and labour standards issues. *The World Economy*, *18*(6), 745–759.

Bodenhausen, G. V., Gabriel, S., & Lineberger, M. (2000). Sadness and susceptibility to judgmental bias: The case of anchoring. *Psychological Science*, *11*(4), 320–323.

Bonan, J., Dagnelie, O., LeMay-Boucher, P., & Tenikue, M. (2012). Is it all about money? A randomized evaluation of the impact of insurance literacy and marketing treatments on the demand for health microinsurance in Senegal. Tech. rep., CEPS/INSTEAD Working Paper No 2012-03.

Bonroy, O., & Constantatos, C. (2008). On the use of labels in credence goods markets. *Journal of Regulatory Economics*, *33*(3), 237–252.

Brown, D. K. (1999). Can Consumer Product Labels Deter Foreign Child Labor Exploitation? Tech. rep., Department of Economics, Tufts University Working paper No.99-19.

Bryan, G. (2013). Ambiguity aversion decreases demand for partial insurance: Evidence from African Farmers. Mimeo, London School of Economics.

Brys, G., Hubert, M., & Struyf, A. (2004a). A Robust Measure of Skewness. *Journal of Computational and Graphical Statistics*, *13*(4), 996–1017.

Brys, G., Hubert, M., & Struyf, A. (2004b). A Robustification of the Jarque-Bera Test of Normality. In J. Antoch (Ed.) *COMPSTAT 2004*, (pp. 753–760). Springer-Verlag.

Brys, G., Hubert, M., & Struyf, A. (2006). Robust Measures of Tail Weight. *Computational Statistics and Data Analysis*, (50), 733–759.

Cai, J., De Janvry, A., & Sadoulet, E. (2015). Social Networks and the Decision to Insure. *American Economic Journal: Applied Economics*, *7*(2), 81–108.

Cai, J., & Song, C. (2013). Do hypothetical experiences affect real financial decisions? Evidence from insurance take-up. Mimeo.

Capuno, J. J., Kraft, A. D., Quimbo, S., Tan Jr, C. R., & Wagstaff, A. (2014). Effects of interventions to raise voluntary enrollment in a social health insurance scheme: a cluster randomized trial. Tech. Rep. 6893, World Bank Policy Research Working Paper.

Carter, M., de Janvry, A., Sadoulet, E., & Sarris, A. (2014). Index-based weather insurance for developing countries: A review of evidence and a set of propositions for up-scaling. Tech. rep., FERDI Working paper, No P111.

Carter, M., Elabed, G., & Serfilippi, E. (2015). Behavioral economic insights on index insurance design. *Agricultural Finance Review*, *75*(1), 8–18.

Carter, M. R., Cheng, L., & Sarris, A. (2011). The impact of interlinked index insurance and credit contracts on financial market deepening and small farm productivity. In *Annual Meeting of the American Applied Economics Association, Pittsburgh PA, July*, (pp. 24–26).

Carter, M. R., Galarza, F., & Boucher, S. (2007). Underwriting area-based yield insurance to crowd-in credit supply and demand. *Savings and Development*, (pp. 335–362).

Cason, T. N., Gangadharan, L., & Maitra, P. (2012). Moral hazard and peer monitoring in a laboratory microfinance experiment. *Journal of Economic Behavior & Organization*, *82*(1), 192 – 209.

Cassar, A., Crowley, L., & Wydick, B. (2007). The effect of social capital on group loan repayment: Evidence from field experiments*. *The Economic Journal*, *117*(517), F85–F106.

Chapman, G. B., & Johnson, E. J. (1994). The limits of anchoring. *Journal of Behavioral Decision Making*, *7*(4), 223–242.

Chowdhury, P. R. (2005). Group-lending: Sequential financing, lender monitoring and joint liability. *Journal of development Economics*, *77*(2), 415–439.

Clarke, D., Das, N., de Nicola, F., Hill, R. V., Kumar, N., & Mehta, P. (2012). The value of (customized) insurance for farmers in rural Bangladesh. In *Research Conference on Microinsurance*.

Clarke, D., & Dercon, S. (2009). Insurance, credit and safety nets for the poor in a world of risk. Tech. rep., United Nations, Department of Economics and Social Affairs, Working Papers.

Clarke, D., & Kalani, G. (2011). Microinsurance decisions: evidence from Ethiopia. Mimeo.

Clarke, D. J. (2011a). A theory of rational demand for index insurance. Tech. rep., University of Oxford, Department of Economics Discussion Paper 572.

Clarke, D. J. (2011b). Reinsuring the poor: Group microinsurance design and costly state verification. Tech. rep., University of Oxford, Department of Economics, Economics Series Working Papers.

Cole, S., Giné, X., Tobacman, J., Townsend, R., Topalova, P., & Vickery, J. (2013). Barriers to household risk management: evidence from India. *American economic journal. Applied economics*, *5*(1), 104–135.

Cole, S., Stein, D., & Tobacman, J. (2014). Dynamics of demand for index insurance: Evidence from a long-run field experiment. *The American Economic Review*, *104*(5), 284–290.

Cranfield, J., Henson, S., Northey, J., & Masakure, O. (2010). An assessment of consumer preference for Fair Trade coffee in Toronto and Vancouver. *Agribusiness*, *26*(2), 307–325.

Crayen, D., Hainz, C., & de MartÍnez, C. S. (2013). Remittances, banking status and the usage of insurance schemes. *The Journal of Development Studies*, *49*(6), 861–875.

Criel, B., & Waelkens, M. P. (2003). Declining subscriptions to the maliando mutual health organisation in Guinea-Conakry (West Africa): What is going wrong? *Social Science & Medicine*, *57*(7), 1205–1219.

Croux, C., & Rousseeuw, P. J. (1992). Time-Efficient Algorithms for Two Highly Robust Estimators of Scale. *Computational Statistics*, (1), 411–428.

Darby, M. R., & Karni, E. (1973). Free competition and the optimal amount of fraud. *Journal of law and economics*, *16*(1), 67–88.

Davies, R. B. (2005). Abstinence from child labor and profit seeking. *Journal of Development Economics*, *76*(1), 251–263.

De Allegri, M., Sanon, M., Bridges, J., & Sauerborn, R. (2006). Understanding consumers' preferences and decision to enrol in community-based health insurance in rural West Africa. *Health policy*, *76*(1), 58–71.

De Bock, O., & Ontiveros, D. (2013). Literature review on the impact of microinsurance. Tech. Rep. 35, Microinsurance Innovation Facility, Research Paper No 35.

De Janvry, A., McIntosh, C., & Sadoulet, E. (2012). Fair Trade and Free Entry: Can a Disequilibrium Market Serve as a Development Tool? *Review of Economics and Statistics*, *97(3)*(00), 567–573.

De Pelsmacker, P., Driesen, L., & Rayp, G. (2005). Do consumers care about ethics? Willingness to pay for fair-trade coffee. *Journal of consumer affairs*, *39*(2), 363–385.

Delavande, A., Giné, X., & McKenzie, D. (2011a). Eliciting probabilistic expectations with visual aids in developing countries: how sensitive are answers to variations in elicitation design? *Journal of Applied Econometrics*, *26*(3), 479–497.

Delavande, A., Giné, X., & McKenzie, D. (2011b). Measuring subjective expectations in developing countries: A critical review and new evidence. *Journal of Development Economics*, *94*(2), 151–163.

Delavande, A., & Kohler, H.-P. (2009). Subjective Expectations in the Context of HIV/AIDS in Malawi. *Demographic research*, *20*, 817.

Dercon, S., Gunning, J. W., & Zeitlin, A. (2011). The demand for insurance under limited credibility: Evidence from Kenya. In *International Development Conference, DIAL.*

Dercon, S., Hill, R. V., Clarke, D., Outes-Leon, I., & Taffesse, A. S. (2014). Offering rainfall insurance to informal insurance groups: Evidence from a field experiment in Ethiopia. *Journal of Development Economics*, *106*, 132–143.

Diamond, P. A., & Hausman, J. A. (1994). Contingent valuation: Is some number better than no number? *The Journal of economic perspectives*, (pp. 45–64).

Doepke, M., & Zilibotti, F. (2010). Do international labor standards contribute to the persistence of the child-labor problem? *Journal of Economic Growth*, *15*(1), 1–31.

Doherty, N. A., & Schlesinger, H. (1990). Rational Insurance Purchasing: Consideration of Contract Nonperformance. *The Quarterly Journal of Economics*, (pp. 243–253).

Dominitz, J., & Manski, C. F. (1997). Using expectations data to study subjective income expectations. *Journal of the American Statistical Association*, *92*(439), 855–867.

Dong, H., De Allegri, M., Gnawali, D., Souares, A., & Sauerborn, R. (2009). Drop-out analysis of community-based health insurance membership at Nouna, Burkina Faso. *Health policy*, *92*(2), 174–179.

Dragusanu, R., & Nunn, N. (2014). The Impacts of Fair Trade Certification: Evidence From Coffee Producers in Costa Rica (Preliminary and Incomplete). Mimeo, Havard University.

Edmonds, E. V. (2007). The Economics of Consumer Actions against Products with Child Labor Content. Tech. rep., prepared for the Child Labor World Atlas, Hugh Hindman, ed. ME Sharpe Publishers, New York.

Ekman, B. (2004). Community-based health insurance in low-income countries: a systematic review of the evidence. *Health policy and planning*, *19*(5), 249–270.

Elabed, G., Bellemare, M. F., Carter, M. R., & Guirkinger, C. (2013). Managing basis risk with multiscale index insurance. *Agricultural Economics*, *44*(4-5), 419–431.

Elabed, G., & Carter, M. R. (2015). Compound-risk aversion, ambiguity and the willingness to pay for microinsurance. *Journal of Economic Behavior and Organization*, *118*, 150–166.

Elfenbein, D. W., & McManus, B. (2010). A greater price for a greater good? Evidence that consumers pay more for charity-linked products. *American Economic Journal: Economic Policy*, *2*(2), 28–60.

Eling, M., Pradhan, S., & Schmit, J. T. (2014). The determinants of microinsurance demand. *The Geneva Papers on Risk and Insurance-Issues and Practice*, *39*(2), 224–263.

Ellsberg, D. (1961). Risk, ambiguity, and the Savage axioms. *The quarterly journal of economics*, (pp. 643–669).

Englich, B., Mussweiler, T., & Strack, F. (2006). Playing dice with criminal sentences: The influence of irrelevant anchors on experts' judicial decision making. *Personality and Social Psychology Bulletin*, *32*(2), 188–200.

Fafchamps, M. (1992). Solidarity networks in preindustrial societies: Rational peasants with a moral economy. *Economic development and cultural change*, (pp. 147–174).

Fafchamps, M., & Lund, S. (2003). Risk-sharing networks in rural Philippines. *Journal of development Economics*, *71*(2), 261–287.

Fischer, G. (2013). Contract structure, risk-sharing, and investment choice. *Econometrica*, *81*(3), 883–939.

Fischer, R., & Serra, P. (2000). Standards and protection. *Journal of International Economics*, *52*(2), 377–400.

Fitzpatrick, A., Magnoni, B., & Thornton, R. L. (2011). Microinsurance utilization in Nicaragua : A report on effects on children, retention, and health. Tech. rep., ILO Microinsurance Innovation Facility Research Paper No. 5.

Freeman, R. B. (1998). What role for labor standards in the global economy? Harvard University and NBER, Centre for Economic Performance, LSE, mimeo.

Fung, A., Sabel, C., & O'Rourke, D. (2001). Stepping up labor standards. *Boston Rev*, *26*(1), 4–20.

Furnham, A., & Boo, H. C. (2011). A literature review of the anchoring effect. *The Journal of Socio-Economics*, *40*(1), 35–42.

Gächter, S., & Fehr, E. (2000). Cooperation and punishment in public goods experiments. *American Economic Review*, *90*(4), 980–994.

Galarza, F., & Carter, M. R. (2011). Risk preferences and demand for insurance in Peru: A field experiment. Tech. rep., Working Paper DD/11/08, Universidad del Pacifico, Peru.

Ganzach, Y., & Karsahi, N. (1995). Message framing and buying behavior: A field experiment. *Journal of Business Research*, *32*(1), 11–17.

Gaurav, S., Cole, S., & Tobacman, J. (2011). Marketing complex financial products in emerging markets: Evidence from rainfall insurance in India. *Journal of marketing research*, *48*(SPL), 150–162.

Ghatak, M., & Guinnane, T. W. (1999). The economics of lending with joint liability: theory and practice. *Journal of Development Economics*, *60*(1), 195–228.

Giesbert, L., Steiner, S., & Bendig, M. (2011). Participation in micro life insurance and the use of other financial services in Ghana. *Journal of Risk and Insurance*, *78*(1), 7–35.

Gilovich, T., Vallone, R., & Tversky, A. (1985). The hot hand in basketball: On the misperception of random sequences. *Cognitive psychology*, *17*(3), 295–314.

Giné, X., Jakiela, P., Karlan, D., & Morduch, J. (2010). Microfinance games. *American Economic Journal: Applied Economics*, *2*(3), 60–95.

Giné, X., Karlan, D., & Ngatia, M. (2014). Social networks, financial literacy and index insurance. In M. Lundberg, & F. Mulaj (Eds.) *Enhancing financial capability and behaviour in low- and middle-income countries*. World Bank, Washington, DC.

Giné, X., Townsend, R., & Vickery, J. (2008). Patterns of rainfall insurance participation in rural India. *The World Bank Economic Review*, *22*(3), 539–566.

Giné, X., & Yang, D. (2009). Insurance, credit, and technology adoption: Field experimental evidencefrom Malawi. *Journal of development Economics*, *89*(1), 1–11.

Green, D., Jacowitz, K. E., Kahneman, D., & McFadden, D. (1998). Referendum contingent valuation, anchoring, and willingness to pay for public goods. *Resource and Energy Economics*, *20*(2), 85–116.

Groeneveld, R. A. (1991). An Influence Function Approach to Describing the Skewness of a Distribution. *The American Statistician*, *45*(2), 97–102.

Hainmueller, J., Hiscox, M. J., & Sequeira, S. (2015). Consumer Demand for Fair Trade: Evidence from a Multistore Field Experiment. *Review of Economics and Statistics*, *97*(2), 242–256.

Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., & Stahel, W. A. (1986). *Robust Statistics: The Approach Based on Influence Functions*. New York: John Wiley.

Hermes, N., Lensink, R., & Mehrteab, H. T. (2005). Peer monitoring, social ties and moral hazard in group lending programs: Evidence from eritrea. *World Development*, *33*(1), 149–169.

Hill, R. V., Hoddinott, J., & Kumar, N. (2013). Adoption of weather-index insurance: learning from willingness to pay among a panel of households in rural Ethiopia. *Agricultural Economics*, *44*(4-5), 385–398.

Hill, R. V., Robles, M., et al. (2011). Flexible insurance for heterogeneous farmers: results from a small scale pilot in Ethiopia. Tech. rep., FPRI discussion paper No. 01092.

Hinkley, D. V. (1975). On Power Transformations to Symmetry. *Biometrika*, *62*(1), 101–111.

Hiscox, M. J., & Smyth, N. (2011). Is there consumer demand for Fair labor standards? Evidence from a field experiment. Mimeo, Harvard University.

Hodges, J. L., & Lehmann, E. L. (1963). Estimates of Location Based on Rank Tests. *The Annals of Mathematical Statistics*, *34*(2), 598–611.

Ito, S., & Kono, H. (2010). Why is the take-up of microinsurance so low? Evidence from a health insurance scheme in India. *The Developing Economies*, *48*(1), 74–101.

Jaffee, D. (2008). Better, but not great: The social and environmental benefits and limitations of Fair Trade for indigenous coffee producers in Oaxaca, Mexico. In R. Ruerd (Ed.) *The Impact of Fair Trade*, (pp. 195–222). Wageningen Academic Publishers.

Jarque, C. M., & Bera, A. K. (1980). Efficient Tests for Normality, Homoscedasticity and Serial Independence of Regression Residuals. *Economics Letters*, *6*(3), 255–259.

Jehu-Appiah, C., Aryeetey, G., Agyepong, I., Spaan, E., & Baltussen, R. (2012). Household perceptions and their implications for enrolment in the National Health Insurance Scheme in Ghana. *Health Policy and Planning*, *27*(3), 222–233.

Jensen, N. D., Barrett, C. B., & Mude, A. (2014a). Basis risk and the welfare gains from index insurance: Evidence from northern Kenya. Tech. rep., MPRA Paper No. 59153.

Jensen, N. D., Mude, A., & Barrett, C. B. (2014b). How basis risk and spatiotemporal adverse selection influence demand for index insurance: Evidence from northern Kenya. Tech. rep., MPRA Paper No. 60452.

Jensen, R. (2010). The (perceived) returns to education and the demand for schooling. *The Quarterly Journal of Economics*, *125*(2), 515–548.

Johnson, D. B., & Mizoguchi, T. (1978). Selecting the Kth Element in $X + Y$ and $X_1 + X_2 + \ldots + X_m$. *SIAM Journal of Computing*, (7), 147–153.

Jowett, M. (2003). Do informal risk sharing networks crowd out public voluntary health insurance? Evidence from Vietnam. *Applied economics*, *35*(10), 1153–1161.

Kahneman, D. (2011). *Thinking, fast and slow*. Macmillan.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica: Journal of the Econometric Society*, (pp. 263–291).

Karlan, D. S. (2007). Social connections and group banking. *The Economic Journal*, *117*(517), F52–F84.

Karlan, D. S., Osei, R., Osei-Akoto, I., & Udry, C. (2014). Agricultural decisions after relaxing credit and risk constraints. *The Quarterly Journal of Economics*, *129*(2), 597–652.

Kolmogorov, A. N. (1933). Sulla determinazione empirica di una legge di distribuzione. *Giornale dell'Istituto Italiano degli Attuari*, *4*, 83–91.

Kunreuther, H. C., Pauly, M. V., & McMorrow, S. (2013). *Insurance and behavioral economics: improving decisions in the most misunderstood industry*. Cambridge: Cambridge University Press.

Leblois, A., & Quirion, P. (2013). Agricultural insurances based on meteorological indices: realizations, methods and research challenges. *Meteorological Applications*, *20*(1), 1–9.

Loureiro, M. L., & Lotade, J. (2005). Do fair trade and eco-labels in coffee wake up the consumer conscience? *Ecological Economics*, *53*(1), 129–138.

Luchini, S., & Watson, V. (2013). Uncertainty and framing in a valuation task. *Journal of Economic Psychology*, *39*, 204–214.

Manski, C. F. (2004). Measuring Expectations. *Econometrica*, *72*(5), 1329–1376.

Maskus, K. E. (1997). Should core labor standards be imposed through international trade policy? Tech. Rep. 1817, Policy Research Working Paper 1817, World Bank, Washington, D.C.

McIntosh, C., Sarris, A., & Papadopoulos, F. (2013). Productivity, credit, risk, and the demand for weather index insurance in smallholder agriculture in Ethiopia. *Agricultural Economics*, *44*(4-5), 399–417.

McKenzie, D., Gibson, J., & Stillman, S. (2013). A land of milk and honey with streets paved with gold: Do emigrants have over-optimistic expectations about incomes abroad? *Journal of Development Economics*, *102*, 116–127.

Miranda, M. J., & Farrin, K. (2012). Index insurance for developing countries. *Applied Economic Perspectives and Policy*, *34*(3), 391–427.

Mobarak, A. M., & Rosenzweig, M. R. (2013). Informal risk sharing, index insurance, and risk taking in developing countries. *The American Economic Review*, *103*(3), 375–380.

Molini, V., Keyzer, M., van den Boom, B., Zant, W., et al. (2008). Creating safety nets through semi-parametric index-based insurance: A simulation for Northern Ghana. *Agricultural Finance Review*, *68*(1), 223–246.

Morduch, J. (2006). Microinsurance: The next revolution? In A. V. Banerjee, R. Benabou, & D. Mookherjee (Eds.) *Understanding Poverty*, chap. 22, (pp. 337–356). Oxford University Press.

Morgan, M. G., & Small, M. (1992). *Uncertainty: a guide to dealing with uncertainty in quantitative risk and policy analysis*. Cambridge University Press.

Murshid, K., Zohir, S. C., Milford, A., & Wiig, A. (2003). *Experience from Bangladesh with ethical trading initiatives*. Chr. Michelsen Institute.

Nelson, P. (1970). Information and consumer behavior. *The Journal of Political Economy*, *78*(2), 311–329.

Nikiforakis, N. (2008). Punishment and counter-punishment in public good games: Can we really govern ourselves? *Journal of Public Economics*, *92*(1), 91–112.

Northcraft, G. B., & Neale, M. A. (1987). Experts, amateurs, and real estate: An anchoring-and-adjustment perspective on property pricing decisions. *Organizational behavior and human decision processes*, *39*(1), 84–97.

Norton, M., Osgood, D., Madajewicz, M., Holthaus, E., Peterson, N., Diro, R., Mullally, C., Teh, T.-L., & Gebremichael, M. (2014). Evidence of demand for index insurance: experimental games and commercial transactions in Ethiopia. *Journal of Development Studies*, *50*(5), 630–648.

O'Conor, R. M., Johannesson, M., & Johansson, P.-O. (1999). Stated preferences, real behaviour and anchoring: some empirical evidence. *Environmental and Resource Economics*, *13*(2), 235–248.

Ones, U., & Putterman, L. (2007). The ecology of collective action: A public goods and sanctions experiment with controlled group formation. *Journal of Economic Behavior & Organization*, *62*(4), 495–521.

Ostrom, E., Walker, J., & Gardner, R. (1992). Covenants with and without a Sword: Self-governance Is Possible. *American Political Science Review*, *86*(02), 404–417.

Panda, P., Chakraborty, A., & Dror, D. M. (2015). Building awareness to health insurance among the target population of community-based health insurance schemes in rural India. *Tropical Medicine and International Health*, *20*(8), 1093–1107.

Patankar, M. (2011). Comprehensive risk cover through remote sensing techniques in agriculture insurance for developing countries: A pilot project. Tech. rep., ILO Microinsurance Innovation Facility Research Paper.

Patt, A., Peterson, N., Carter, M., Velez, M., Hess, U., & Suarez, P. (2009). Making index insurance attractive to farmers. *Mitigation and Adaptation Strategies for Global Change*, *14*(8), 737–753.

Patt, A., Suarez, P., & Hess, U. (2010). How do small-holder farmers understand insurance, and how much do they want it? Evidence from Africa. *Global Environmental Change*, *20*(1), 153–161.

Platteau, J.-P. (1991). Traditional systems of social security and hunger insurance: Past achievements and modern challenges. In E. Ahmad, J. Drèze, J. R. Hills, & A. K. Sen (Eds.) *Social security in developing countries*, (pp. 112–170). Oxford: Clarendon Press.

Platteau, J.-P. (1997). Mutual insurance as an elusive concept in traditional rural communities. *The Journal of Development Studies*, *33*(6), 764–796.

Platteau, J.-P., & Abraham, A. (1987). An inquiry into quasi-credit contracts: The role of reciprocal credit and interlinked deals in small-scale fishing communities. *The Journal of Development Studies*, *23*(4), 461–490.

Platteau, J.-P., & Ontiveros, D. U. (2013). Understanding and information failures: Lessons from a health microinsurance program in India. Tech. Rep. 29, ILO Microinsurance Innovation Facility Paper.

Podhorsky, A. (2015). A positive analysis of Fairtrade certification. *Journal of Development Economics*, *116*, 169–185.

Poelman, A., Mojet, J., Lyon, D., & Sefa-Dedeh, S. (2008). The influence of information about organic production and fair trade on preferences for and perception of pineapple. *Food Quality and Preference*, *19*(1), 114–121.

Prasad, M., Kimeldorf, H., Meyer, R., & Robinson, I. (2004). Consumers of the World Unite A Market-based Response to Sweatshops. *Labor Studies Journal*, *29*(3), 57–79.

Rodrik, D. (1996). Labor standards in international trade: do they matter and what do we do about them? In R. Z. Lawrence, D. Rodrik, & J. Whalley (Eds.) *Emerging Agenda for Global Trade: High Stakes for Developing Countries.*. Washington, D.C.: Overseas Development Council.

Roe, B., & Sheldon, I. (2007). Credence good labeling: The efficiency and distributional implications of several policy approaches. *American Journal of Agricultural Economics*, *89*(4), 1020–1033.

Rogers, E. M. (1995). *Diffusion of innovations*. Free Press.

Rousseeuw, P. J., & Croux, C. (1993). Alternatives to the Median Absolute Deviation. *Journal of the American Statistical Association*, *88*(424), 1273–1283.

Ruppert, D. (1987). What is Kurtosis? An Influence Function Approach. *The American Statistician*, *41*(1), 1–5.

Schneider, P., & Diop, F. (2004). Community-based health insurance in Rwanda. *Health financing for poor people-Resource mobilization and risk sharing, Washington DC: World Bank*, (pp. 251–274).

Schultz, E., Metcalfe, M., Gray, B., Dunford, C., Guiteras, R., Kazianga, H., & Szott, A. (2013). The impact of health insurance education on enrollment of microfinance institution clients in the Ghana National Health Insurance scheme, Northern region of Ghana. Tech. Rep. 33, ILO Microinsurance Innovation Facility Research Paper No. 33.

Serfilippi, E., Carter, M., & Guirkinger, C. (2015). Certain and Uncertain Utility and Insurance Demand: Results From a Framed Field Experiment in Burkina Faso. University of Namur, Mimeo.

Sirieix, L., Delanchy, M., Remaud, H., Zepeda, L., & Gurviez, P. (2013). Consumers' perceptions of individual and combined sustainable food labels: a UK pilot investigation. *International Journal of Consumer Studies*, *37*(2), 143–151.

Smirnov, N. V. (1933). Estimate of deviation between empirical distribution functions in two independent samples. *Bulletin Moscow University*, *2*, 3–16.

Solon, G., Haider, S. J., & Wooldridge, J. M. (2015). What are we weighting for? *Journal of Human Resources*, *50*(2), 301–316.

Stein, D. (2014). Dynamics of demand for rainfall index insurance: evidence from a commercial product in India. Tech. Rep. 7035, World Bank Policy Research Working Paper no 7035.

Stein, D., & Tobacman, J. (2012). Weather insured savings accounts. Tech. rep., ILO Microinsurance Innovation Facility Research Paper No. 17.

Stiglitz, J. E. (1990). Peer Monitoring and Credit Markets. *World Bank Economic Review*, *4*(3), 351–66.

Tagbata, D., & Sirieix, L. (2008). Measuring consumer's willingness to pay for organic and Fair Trade products. *International Journal of Consumer Studies*, *32*(5), 479–490.

Thisse, J.-F., & Toulemonde, E. (2010). The distribution of earnings under monopsonistic/polistic competition. Tech. rep., IZA, Discussion Paper No. 5136.

Thornton, R., Hatt, L., Field, E., Islam, M., Diaz, F., & González, M. (2010). Social security health insurance for the informal sector in Nicaragua: a randomized evaluation. *Health economics*, *19*, 181–206.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *science*, *185*(4157), 1124–1131.

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and uncertainty*, *5*(4), 297–323.

Udry, C. (1990). Credit markets in Northern Nigeria: Credit as insurance in a rural economy. *The World Bank Economic Review*, *4*(3), 251–269.

Valkila, J., & Nygren, A. (2010). Impacts of Fair Trade certification on coffee farmers, cooperatives, and laborers in Nicaragua. *Agriculture and Human Values*, *27*(3), 321–333.

Van Exel, N., Brouwer, W., van den Berg, B., & Koopmanschap, M. (2006). With a little help from an anchor: Discussion and evidence of anchoring effects in contingent valuation. *The journal of socio-economics*, *35*(5), 836–853.

Vanderwalle, L. (2015). The Role of Accountants in Indian Self-Help Groups: A Trade-off between Financial and Non-Financial Benefits. Mimeo, The Graduate Institute Geneva.

Wakker, P., Thaler, R., & Tversky, A. (1997). Probabilistic insurance. *Journal of Risk and Uncertainty*, *15*(1), 7–28.

Wilson, T. D., Houston, C. E., Etling, K. M., & Brekke, N. (1996). A new look at anchoring effects: basic anchoring and its antecedents. *Journal of Experimental Psychology: General*, *125*(4), 387.

Wydick, B. (1999). Can social cohesion be harnessed to repair market failures? evidence from group lending in guatemala. *Economic Journal*, (pp. 463–475).

Zago, A. M., & Pick, D. (2004). Labeling policies in food markets: Private incentives, public intervention, and welfare effects. *Journal of Agricultural and Resource Economics*, *1*, 150–165.